

PCT

THE BRITISH LIBRARY
SCIENCE WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/12, 15/54, 9/12 // C12Q 1/68, C12N 15/11	A1	(11) International Publication Number: WO 93/15201 (43) International Publication Date: 5 August 1993 (05.08.93)
---	----	--

(21) International Application Number: PCT/US93/00586

(22) International Filing Date: 22 January 1993 (22.01.93)

(30) Priority data:
826,935 22 January 1992 (22.01.92) US

(60) Parent Application or Grant

(63) Related by Continuation

US
Filed on826,935 (CIP)
22 January 1992 (22.01.92)

(71) Applicant (for all designated States except US): NEW ENGLAND DEACONESS HOSPITAL [US/US]; 185 Pilgrim Road, Boston, MA 02215 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): AVRAHAM, Hava [IL/US]; 50 Radmor Road, Brighton, MA 02135 (US). GROOPMAN, Jerome [US/US]; 79 Druce Street, Brookline, MA 02146 (US). COWLEY, Sally [GB/GB]; 28 Avonmore Road, London W14 8RS (GB). SCADDEN, David [US/US]; 62 Lexington Street, Weston, MA 02193 (US).

(74) Agents: GRANAHAN, Patricia et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US).

(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published

With international search report.
With amended claims.Date of publication of the amended claims:
30 September 1993 (30.09.93)

(54) Title: NOVEL PROTEIN TYROSINE KINASES

(57) Abstract

The identification and isolation of novel protein tyrosine kinase genes present on human megakaryocytic and lymphocytic cells, the proteins encoded by these genes, antibodies specific for the encoded proteins, RNA nucleic acid sequences which hybridize to the genes and methods of use therefor.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SK	Slovak Republic
CI	Côte d'Ivoire	LJ	Liechtenstein	SN	Senegal
CM	Cameroon	LK	Sri Lanka	SU	Soviet Union
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	MC	Monaco	TG	Togo
DE	Germany	MG	Madagascar	UA	Ukraine
DK	Denmark	ML	Mali	US	United States of America
ES	Spain	MN	Mongolia	VN	Viet Nam
FI	Finland				

NOVEL PROTEIN TYROSINE KINASESDescriptionBackground of the Invention

Transduction of signals that regulate cell growth and
5 differentiation is regulated in part by phosphorylation of
various cellular proteins. Protein tyrosine kinases are
enzymes that catalyze this process. Moreover, many act as
growth factor receptors.

Summary of the Invention

10 The present invention relates to novel protein
tyrosine kinase genes present in human megakaryocytic and
lymphocytic cells, the proteins encoded by these genes,
antibodies specific for the encoded proteins, RNA nucleic
acid sequences which hybridize to the genes and methods of
15 use therefor.

The genes isolated as described herein are referred
to, collectively, as protein tyrosine kinase (pTK) genes.
The nucleic acid sequences of these genes, isolated as
discussed herein, show significant homology with
20 previously identified protein tyrosine kinases containing
extracellular domains which function as growth factor
receptors. The pTK genes have been shown to be present in
both megakaryocytic and lymphocytic cells.

The pTK genes of the present invention show
25 significant sequence homology with members of the c-kit
subgroup of growth factor receptors with protein tyrosine
kinase activity. The c-kit subgroup of receptor tyrosine
kinases catalyze the phosphorylation of exogenous
substrates, as well as tyrosine residues within their own
30 polypeptide chains. (Ullrich, A. and Schlessinger, J.,
Cell, 61:203 (1990)). Members of the c-kit subgroup
include FLT/FLK (Fetal Liver Kinase), FGF (Fibroblast

-2-

Growth Factor Receptor) and NGF (Nerve Growth Factor Receptor).

In particular, fourteen pTK genes have been identified. Two pTK genes, referred to as SAL-S1 and SAL-D4 (also referred to as megakaryocyte derived FGF-like receptor tyrosine kinase) were identified in megakaryocytic cells. Five pTK genes, referred to as LpTKs, were identified in lymphocytic cells and have been shown to be present in megakaryocytes as well. One pTK gene, referred to as HpTKs, was identified in human hepatoma cells. Six pTK genes, referred to as bpTK genes, found in human brain tissue.

SAL-S1 is related to the FLT/FLK family of pTKs. SAL-D4 is related to the FGF receptor family of pTKs, and one LpTK (LpTK 3) is related to the NGF receptor family of pTKs.

The pTK genes, which are the subject of the present invention, were identified using two sets of degenerative oligonucleotide primers: a first set which amplifies all pTK DNA segments (SEQ ID NOS:1-2), and a second set which amplifies highly conserved sequences present in the catalytic domain of the c-kit subgroup of pTKs (SEQ ID NOS:3-4). The pTK genes identified in this manner are described below.

SAL-S1 is expressed in several megakaryocytic cell lines, but not in erythroid cell lines. The nucleotide sequence of SAL-S1 was obtained, revealing a sequence containing 158 base pairs. (SEQ ID NO:5). This isolated DNA fragment encoded an amino acid sequence (SEQ ID NO:6) which exhibited significant sequence homology with known protein tyrosine kinases of the FLT/FLK family. The full

-3-

length gene sequence (SEQ ID NO: 17) contains 6827 b.p. and the deduced amino acid sequence (SEQ ID NO: 18) contains 349 residues.

SAL-D4, also expressed in megakaryocytic cells, is a DNA fragment containing the nucleotide sequence of 141 base pairs. (SEQ ID NO:7). This isolated DNA fragment encoded an amino acid sequence (SEQ ID NO:8) which exhibited significant sequence homology with known protein tyrosine kinases of the FGF receptor family.

10 The LpTKs, including LpTK 2, LpTK 3, LpTK 4, and LpTK 13 and LpTK 25, are expressed in lymphocytic cells, as well as megakaryocytic cells. The nucleotide sequence (151 base pairs) of the LpTK 3 gene was obtained (SEQ ID NO:11), and exhibited significant homology with known protein tyrosine kinases of the NGF receptor family. The nucleotide sequences of the LpTK 2, LpTK 4, and LpTK 13 genes contained 149 base pairs (SEQ ID NO:9), 137 base pairs (SEQ ID NO:13), and 211 base pairs (SEQ ID NO:15) respectively. LpTK 25 has a nucleotide sequence of 3120 b.p. (SEQ ID NO: 22). A full length gene sequence has been obtained for LpTK 2 (SEQ ID NO: 19) which contains 7606 b.p. Additional sequencing of LpTK 4 revealed a sequence of 404 b.p. (SEQ ID NO: 21).

25 The HpTK 5 gene, expressed in human hepatoma cells, has a nucleotide sequence of 3120 b.p. (SEQ ID NO: 22). Nucleotide sequences of the bpTK's, including bpTK 1, bpTK 2, bpTK 3, bpTK 4, bpTK 5 and bpTK 7 are expressed in human brain tissue encode proteins having the amino acid sequences of SEQ ID NOS: 25-30 respectively.

30 Thus the present invention includes DNA isolated from a human megakaryocytic cell line, which hybridizes to a

-4-

DNA fragment which hybridizes to DNA encoding an amino acid sequence which is highly conserved in the catalytic domain of protein tyrosine kinases of the c-kit subgroup.

The present invention also includes the proteins
5 encoded by the pTK genes identified as described herein, which exhibit significant sequence homology with members of the c-kit subgroup of pTKs (i.e. FLT/FLK (SAL-S1), FGF receptor (SAL-D4) or NGF receptor (LpTKs)) as well as the proteins encoded by HpTK 5 and the bpTKs. The present
10 invention also includes SAL-S1, SAL-D4, and LpTK, HpTK and bpTK homologues or equivalents (i.e., proteins which have amino acid sequences substantially similar, but not identical, to that of SAL-S1, SAL-D4, the LpTKs HpTK and the bpTKs, which exhibit tyrosine kinase activity.) This
15 invention further includes peptides (SAL-S1, SAL-D4, LpTK, HpTK and bpTK fragments) which retain tyrosine kinase activity, yet are less than the entire SAL-S1, SAL-D4, LpTK, HpTK and bpTK sequences), monoclonal and polyclonal antibodies specific for SAL-S1, SAL-D4, the LpTKs, HpTK
20 and the bpTKs, and uses for the SAL-S1, SAL-D4, the LpTK, HpTK and the bpTK nucleic acid sequences and SAL-S1, SAL-D4, LpTK, HpTK and bpTK equivalents.

The present invention further includes nucleic acid sequences which hybridize with DNA or RNA encoding the
25 proteins described herein, which exhibit significant sequence homology with the FLT/FLK, FGF receptor or NGF receptor family of protein tyrosine kinases contained within the c-kit subgroup. Such nucleic acid sequences are useful as probes to identify pTK genes in other
30 vertebrates, particularly mammals, and in other cell types. They can also be used as anti-sense

-5-

oligonucleotides to inhibit protein tyrosine kinase activity, both in vitro and in vivo.

The SAL-S1, SAL-D4, LpTK, HpTK and bpTK, tyrosine kinases of the present invention can be used as target proteins in conjunction with the development of drugs and therapeutics to modulate cell growth, differentiation and other metabolic functions. The SAL-S1, SAL-D4, LpTK, HpTK or bpTK proteins can be used as agonists or antagonists to other tyrosine kinases. The SAL-S1, SAL-D4, LpTK, HpTK or bpTK tyrosine kinases can also be instrumental in the modulation of megakaryocyte and/or platelet adhesion interactions.

In addition, the SAL-S1, SAL-D4, LpTK, HpTK and bpTK tyrosine kinases can be used in screening assays to detect cellular growth and/or differentiation factors. Using standard laboratory techniques, the ligands of the pTKs of the present invention can be identified. Once identified, assays can be designed to detect these ligands present endogenously, within cells, as well as exogenously, in extra cellular fluids. Assays can also be designed as diagnostic aids to detect these ligands in body fluids such as blood and urine.

Brief Description of the Drawings

Figure 1 depicts the nucleotide sequence of SAL-S1 (SEQ ID NO: 5) and the deduced amino acid sequence (SEQ ID NO:6).

Figure 2 depicts the nucleotide sequence of SAL-D4 (SEQ ID NO:7) and its deduced amino acid sequence (SEQ ID NO:8).

-6-

Figure 3A depicts the nucleotide sequence (SEQ ID NO:9) and its deduced amino acid sequence (SEQ ID NO:10) for LpTK 2.

Figure 3B depicts the nucleotide sequence (SEQ ID NO:11) and its deduced amino acid sequence (SEQ ID NO:12) for LpTK 3.

Figure 3C depicts the nucleotide sequence (SEQ ID NO:13) and its deduced amino acid sequence (SEQ ID NO:14) for LpTK 4.

Figure 3D depicts the nucleotide sequence (SEQ ID NO:15) and its deduced amino acid sequence (SEQ ID NO:16) for the LpTK 13.

Figure 4A-4J depicts the full-length nucleotide sequence (SEQ ID NO: 17) and its deduced amino acid sequence (SEQ ID NO: 18) for SAL-S1.

Figure 5A-5J depicts the full length nucleotide sequence (SEQ ID NO: 19) and the deduced amino acid sequence (SEQ ID NO: 20) for LpTK2.

Figure 6 depicts the partial nucleotide sequence (SEQ ID NO: 21) for LpTK4.

Figure 7A-7D depicts the full length nucleotide sequence (SEQ ID NO: 22) for LpTK25.

Figure 8A-8F depicts the full length nucleotide sequence (SEQ ID NO: 23) and the deduced amino acid sequence (SEQ ID NO: 24) for HpTK5.

Figure 9 depicts the amino acid sequence (SEQ ID NO: 25) of bpTK1.

Figure 10 depicts the amino acid sequence (SEQ ID NO: 26) of bpTK2.

Figure 11 depicts the amino acid sequence (SEQ ID NO: 27) of bpTK3.

-7-

Figure 12 depicts the amino acid sequence (SEQ ID NO: 28) of bpTK4.

Figure 13 depicts the amino acid sequence (SEQ ID NO: 29) of bpTK5.

5 Figure 14 depicts the amino acid sequence (SEQ ID NO: 30) of bpTK7.

Detailed Description of the Invention

Novel protein tyrosine kinase genes have been identified, their nucleic acid sequences determined, and
10 the amino acid sequences of the encoded proteins deduced. The genes isolated as described herein are referred to, collectively, as protein tyrosine kinase (pTK) genes. The nucleic acid sequences of these genes, isolated as discussed herein, show significant homology to
15 with previously identified protein tyrosine kinases containing extracellular domains which function as growth factor receptors. These genes have been shown to be present in both megakaryocytic and lymphocytic cells.

To facilitate the isolation and identification of
20 these novel pTKs, two sets of DNA probes were used, as described in the Exemplification. The first set consisted of two degenerative oligonucleotide sequences, pTK 1 (SEQ ID NO:1) and pTK 2 (SEQ ID NO:2) (Matthews, W. Cell 65: 1143 (1991; Wilks, A. F. Proc. Natl. Acad. Sci. USA
25 86:1603 (1989)). These sequences were used as primers in a polymerase chain reaction to amplify tyrosine kinase DNA segments. (Mullis, K. et al., Cold Spring Harbor Symp. Advan. Biol. 51:263 (1986).

The second set consisted of two oligonucleotide
30 sequences, pTK 3 (SEQ ID NO:3) and pTKKW (SEQ ID NO:4)

-8-

designed to amplify the nucleic acid sequence which encodes the highly conserved regions of the catalytic domains of the c-kit family of protein tyrosine kinases. These sequences were used as primers in the polymerase chain reaction in a second round of DNA amplification. Using this two-step amplification procedure, DNA fragments which hybridized to these pTK primers were identified, isolated and subsequently sequenced.

In particular, fourteen pTK genes exhibiting significant homology with the c-kit subgroup of protein tyrosine kinases have been identified. Two pTK genes, referred to as SAL-S1 and SAL-D4 (also referred to as megakaryocyte derived FGF-like receptor) were identified in several megakaryocytic cell lines, including CMK 11-5, DAMI, UT-7 and UT-7 grown in erythropoietin, but not in the erythroid cell lines HEL, PMA stimulated HEL cells, or K562. Five pTK genes, referred to as LpTKs, were identified in lymphocytic, as well as in megakaryocytic cells. One pTK gene, referred to as HpTK5 was identified in human hepatoma cells and six genes, referred to as bpTKs, were identified in human brain tissue.

SAL-S1 (SEQ ID NO:6 and 18) encoded by the nucleic acid sequence SEQ ID NOS:5 and 17, exhibits significant homology with the FLT/FLK family of pTKs. SAL-D4 (SEQ ID NO:8) encoded by SEQ ID NO:7, is related to the FGF receptor family of pTKs, and one LpTK (LpTK 3 (SEQ ID NO:12) encoded by the SEQ ID NO:11, is related to the NGF receptor family of pTKs. The remaining LpTKs, LpTK2 (SEQ ID NO:10) encoded by SEQ ID NO:9; LpTK4 (SEQ ID NO:14) encoded by SEQ ID NO:13; LpTK13 (SEQ ID NO:16) encoded by SEQ ID NO:15 LpTK25 encoded by SEQ ID NO: 22, also exhibit

-9-

sequence homology with known protein tyrosine kinases
(Data not shown).

HpTK5 (SEQ ID NO: 24) encoded by SEQ ID NO: 23 and
the bpTKs 1, 2, 3, 4, 5 and 7 (SEQ ID NOS: 25-30
5 respectively, also exhibit sequence homology with known
protein tyrosine kinases.

Thus, as described above, DNA which hybridize with
DNA encoding amino acid sequences present in the catalytic
domain of a protein tyrosine kinase of the c-kit subgroup
10 of protein kinases have been isolated and sequenced.
These isolated DNA sequences, collectively referred to as
pTKs genes, (and their deduced amino acid sequences) have
been shown to exhibit significant sequence homology with
known members of receptor tyrosine kinase families.

15 Once isolated, these DNA fragments can be amplified
using known standard techniques such as PCR. These
amplified fragments can then be cloned into appropriate
cloning vectors and their DNA sequences determined.

These DNA sequences can be excised from the cloning
20 vectors, labeled with a radiolabeled nucleotide such as
 ^{32}P and used to screen appropriate cDNA libraries to
obtain the full-length cDNA clone.

The pTk genes as described above have been isolated
from the source in which they occur naturally, i.e.
25 megakaryocyte and lymphocytic cells. The present invention
is intended to include pTk genes produced using genetic
engineering techniques, such as recombinant technology, as
well as pTk genes that are synthesized chemically.

The deduced amino acid sequences of the pTK genes
30 include amino acid sequences which encode peptides
exhibiting significant homology with the catalytic domain

-10-

of protein tyrosine kinases of the c-kit subgroup of tyrosine kinases. These proteins, encoded by the pTk genes, can include sequences in which functionally equivalent amino acid residues are substituted for
5 residues within the sequence, resulting in a silent change, that is a change not detected phenotypically. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent,
10 resulting in a silent substitution.

In addition, the protein structure can be modified by deletions, additions, inversion, insertions or substitutions of one or more amino acid residues in the sequence which do not substantially detract from the
15 desired functional tyrosine kinases properties of the peptide.

Modified pTKs of the present invention, with receptor tyrosine kinase activity can be made using recombinant DNA techniques, such as excising it from a vector containing a
20 cDNA encoding such a protein, or by synthesizing DNA encoding the desired protein mechanically and/or chemically using known techniques.

An alternate approach to producing the pTKs of the present invention is to use peptide synthesis to make a
25 peptide or polypeptide having the amino acid sequence of such a protein. The peptides or modified equivalents thereof, can be synthesized directly by standard solid or liquid phase chemistries for peptide synthesis.

Preferably, the pTKs of the present invention will be
30 produced by inserting DNA encoding the proteins into an appropriate vector/host system where it will be expressed.

-11-

The DNA sequences can be obtained from sources in which they occur naturally, can be chemically synthesized or can be produced using standard recombinant technology.

This invention also pertains to an expression vector
5 comprising a pTK gene of the present invention, encoding for a protein which exhibits receptor tyrosine kinase activity.

The pTK genes of the present invention can be used for a number of diagnostic and therapeutic purposes. For
10 example, the nucleic acid sequences of the pTK genes can be used as probes to identify other protein tyrosine kinases present in other cell types, including eukaryotic and prokaryotic cell types.

The nucleic acid sequences can be used to design
15 drugs that directly inhibit the kinase activity of protein tyrosine kinases, or to design peptides that bind to the catalytic domain of tyrosine kinases, thus inhibiting their activity. These sequences can also be used to design anti-sense nucleotides that can also inhibit, or
20 destroy, tyrosine kinase activity. Such inhibition of tyrosine kinase activity would be desirable in pathological states where decreased cellular proliferation would be beneficial, such as leukemias or other malignancies.

25 The nucleic acid sequences can also be used to design drugs, peptides or anti-sense nucleotides as above, but with enhancing, rather than inhibitory effects, on tyrosine kinases. Such enhanced tyrosine kinase activity would result in increasing the phosphorylation of
30 substrates (exogenous, as well as endogenous tyrosine residues). Enhanced effects would be desirable in states

-12-

where increased cellular proliferation would be beneficial, such as anemias, bleeding disorders and during surgical procedures.

The pTK genes of the present invention can also be
5 used to obtain soluble fragments of receptor tyrosine kinases, capable of binding their respective ligands (i.e. fibroblast growth factor).

pTK genes encoding soluble receptor tyrosine kinase fragments can be produced using recombinant DNA techniques
10 or synthetically. In either case, the DNA obtained encodes a soluble pTK fragment which lacks a substantial portion of the hydrophobic transmembrane region to permit solubilization of the fragment.

These soluble pTK protein fragments can be introduced
15 exogenously to act as competitors with the endogenous, membrane bound pTK for their respective ligands, thus inhibiting tyrosine kinase activity. Alternately, a modified soluble pTK protein fragment can be introduced which binds the ligand but does not activate kinase
20 activity.

These soluble pTK protein fragments can also be used in binding assays to detect ligands such as growth and differentiation factors. Once these ligands are identified, they may be altered or modified to inhibit or
25 enhance kinase activity. For example, the ligands may be modified or attached to substances that are toxic to the cell, such a ricin, thus destroying the target cell. The substance may be a super-activating substance which, after binding to the pTK, may substantially increase the kinase
30 activity, or activate other growth factors.

-13-

pTk genes of the present invention would also be useful to develop diagnostic tools for in vitro screening assays for ligands such as growth factors or differentiation factors that inhibit or enhance kinase activity. The proteins encoded by the pTk genes can also be used in such assays, or as immunogens to produce monoclonal or polyclonal antibodies to be used in such assays.

Such antibodies can also be used in methods of treating conditions in which an individual would benefit therapeutically if protein tyrosine kinase activity could be modified, such as increasing platelet production in bleeding disorders.

The present invention will now be illustrated by the following Exemplification, which is not intended to be limiting in any way.

Exemplification: The Identification and Isolation of the pTk Genes

To facilitate the isolation and identification of these novel pTk genes, two sets of DNA probes were used. (See Table).

The first set consisted of two degenerative oligonucleotide sequences, pTK 1 (SEQ ID NO:1) and pTK 2 (SEQ ID NO:2). These sequences were used as polymerase chain reaction (PCR) primers, using standard PCR techniques, to amplify tyrosine kinase DNA segments.

The second set consisted of two oligonucleotide sequences, pTK 3 (SEQ ID NO:3) and pTKKW (SEQ ID NO:4) selected from the highly conserved regions of the catalytic domains of the c-kit subgroup of protein

-14-

tyrosine kinases. These sequences were also used as polymerase chain reaction primers in a second round of DNA amplification. Using this two-step amplification procedure, DNA fragments which hybridized to these pTK
5 primers were identified, isolated and subsequently sequenced using known laboratory techniques.

TABLEFirst Round of Amplification

PTK1

10 CGGATCCACAGNGACCT

PTK2

GGAATTCCAAAGGACCAGACGTC

Second Round of Amplification

PTK3 (kit family specific)

15 CGGATCCATCCACAGAGATGT

PTKKW (kit family specific)

GGAATTCCTTCAGGAGCCATCCACTT

Equivalents

Those skilled in the art will recognize, or be able
20 to ascertain using no more than routine experimentation,
many equivalents to the specific embodiments of the
invention described herein. Such equivalents are intended
to be encompassed by the following claims.

-15-

CLAIMS

The invention claimed is:

1. Isolated DNA of human megakaryocytic origin which hybridizes to a DNA fragment which hybridizes to DNA
5 encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein tyrosine kinases.
2. Isolated DNA of Claim 1 having a nucleotide sequence selected from the group of nucleotide sequences
10 consisting of:
 - a) SAL-S1 (SEQ ID NOS:5 and 7);
 - b) SAL-D4 (SEQ ID NO:7);
 - c) LpTK 2 (SEQ ID NOS:9 and 19);
 - d) LpTK 3 (SEQ ID NO:11);
 - 15 e) LpTK 4 (SEQ ID NOS:13 and 21);
 - f) LpTK 13 (SEQ ID NO:15);
 - g) LpTK 25 (SEQ ID NO: 22 and
 - h) HpTK 5 (SEQ ID NO: 23).
3. Isolated DNA of Claim 1 which encodes an amino acid
20 sequence selected from the group consisting of:
 - a) SAL-S1 (SEQ ID NOS:6 and 18);
 - b) SAL-D4 (SEQ ID NO:8);
 - c) LpTK 2 (SEQ ID NOS:10 and 20);
 - d) LpTK 3 (SEQ ID NO:12);
 - 25 e) LpTK 4 (SEQ ID NO:14);
 - f) LpTK 13 (SEQ ID NO:16);
 - g) HpTK 5 (SEQ ID NO:24);

-16-

- h) bpTK 1 (SEQ ID NO:25);
- i) bpTK 2 (SEQ ID NO:26);
- j) bpTK 3 (SEQ ID NO:27);
- k) bpTK 4 (SEQ ID NO:28);
- 5 l) bpTK 5 (SEQ ID NO:29); and
- m) bpTK 7 (SEQ ID NO:30).

4. Isolated DNA of human megakaryocytic origin which comprises a DNA fragment whose sequence encodes an amino acid sequence present in the catalytic domain
10 of a protein tyrosine kinase of the c-kit subgroup of protein tyrosine kinases.

5. Isolated DNA of Claim 4 which encodes an amino acid sequence selected from the group consisting of:
- a) SAL-S1 (SEQ ID NO:6);
 - 15 b) SAL-D4 (SEQ ID NO:8);
 - c) LpTK 2 (SEQ ID NO:10);
 - d) LpTK 3 (SEQ ID NO:12);
 - e) LpTK 4 (SEQ ID NO:14); and
 - f) LpTK 13 (SEQ ID NO:16).
 - 20 g) HpTK 5 (SEQ ID NO:24);
 - h) bpTK 1 (SEQ ID NO:25);
 - i) bpTK 2 (SEQ ID NO:26);
 - j) bpTK 3 (SEQ ID NO:27);
 - k) bpTK 4 (SEQ ID NO:28);
 - 25 l) bpTK 5 (SEQ ID NO:29); and
 - m) bpTK 7 (SEQ ID NO:30).

-17-

6. A homogeneous protein of human megakaryocytic origin which includes an amino acid sequence exhibiting sequence homology with the catalytic domain of tyrosine kinases of the c-kit family.
- 5 7. A homogeneous protein of Claim 6 in which the amino acid sequence is selected from the group consisting of:
 - a) SAL-S1 (SEQ ID NO:6);
 - b) SAL-D4 (SEQ ID NO:8);
 - 10 c) LpTK 2 (SEQ ID NO:10);
 - d) LpTK 3 (SEQ ID NO:12);
 - e) LpTK 4 (SEQ ID NO:14); and
 - f) LpTK 13 (SEQ ID NO:16).
 - g) HpTK 5 (SEQ ID NO:24);
 - 15 h) bpTK 1 (SEQ ID NO:25);
 - i) bpTK 2 (SEQ ID NO:26);
 - j) bpTK 3 (SEQ ID NO:27);
 - k) bpTK 4 (SEQ ID NO:28);
 - l) bpTK 5 (SEQ ID NO:29); and
 - 20 m) bpTK 7 (SEQ ID NO:30).
8. A protein of human megakaryocytic origin which exhibits significant sequence homology with the FLT/FLK family of protein tyrosine kinases.
9. A protein of Claim 8 encoded by the nucleotide
25 sequence (SEQ ID NO:5).
10. A protein of Claim 8 encoded by the amino acid sequence (SEQ ID NO:6).

-19-

19. A DNA expression vector containing a DNA sequence of human megakaryocytic origin which hybridizes to a DNA fragment which hybridizes to DNA encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein kinases.
20. The DNA expression vector of Claim 17 containing a DNA sequence selected from the group consisting of:
- a) SAL-S1 (SEQ ID NO:5);
 - b) SAL-D4 (SEQ ID NO:7);
 - c) LpTK 2 (SEQ ID NO:9);
 - d) LpTK 3 (SEQ ID NO:11);
 - e) LpTK 4 (SEQ ID NO:13); and
 - f) LpTK 13 (SEQ ID NO:15).
 - g) LpTK 25 (SEQ ID NO: 22 and
 - h) HpTK 5 (SEQ ID NO: 23).
21. A cell transformed by the expression vector of Claim 17.

AMENDED CLAIMS

[received by the International Bureau on 31 August 1993 (31.08.93);
original claims 1, 2, 4-7 and 14-21 amended; remaining claims unchanged (5 pages)]

1. Isolated DNA of human origin which hybridizes to a
DNA fragment which hybridizes to DNA encoding an
amino acid sequence present in the catalytic domain
of a protein tyrosine kinase of the c-kit subgroup of
protein tyrosine kinases.
2. Isolated DNA of Claim 1 having a nucleotide sequence
selected from the group of nucleotide sequences
consisting of:
 - a) SAL-S1 (SEQ ID NOS:5 and 17);
 - b) SAL-D4 (SEQ ID NO:7);
 - c) LpTK 2 (SEQ ID NOS:9 and 19);
 - d) LpTK 3 (SEQ ID NO:11);
 - e) LpTK 4 (SEQ ID NOS:13 and 21);
 - f) LpTK 13 (SEQ ID NO:15);
 - g) LpTK 25 (SEQ ID NO: 22); and
 - h) HpTK 5 (SEQ ID NO: 23).
3. Isolated DNA of Claim 1 which encodes an amino acid
sequence selected from the group consisting of:
 - a) SAL-S1 (SEQ ID NOS:6 and 18);
 - b) SAL-D4 (SEQ ID NO:8);
 - c) LpTK 2 (SEQ ID NOS:10 and 20);
 - d) LpTK 3 (SEQ ID NO:12);
 - e) LpTK 4 (SEQ ID NO:14);
 - f) LpTK 13 (SEQ ID NO:16);
 - g) HpTK 5 (SEQ ID NO:24);

- h) bpTK 1 (SEQ ID NO:25);
i) bpTK 2 (SEQ ID NO:26);
j) bpTK 3 (SEQ ID NO:27);
k) bpTK 4 (SEQ ID NO:28);
5 l) bpTK 5 (SEQ ID NO:29); and
m) bpTK 7 (SEQ ID NO:30).
4. Isolated DNA of human origin which comprises a DNA
fragment whose sequence encodes an amino acid
sequence present in the catalytic domain of a protein
10 tyrosine kinase of the c-kit subgroup of protein
tyrosine kinases.
5. Isolated DNA of Claim 4 which encodes an amino acid
sequence selected from the group consisting of:
15 a) SAL-S1 (SEQ ID NOS:6 and 18);
b) SAL-D4 (SEQ ID NO:8);
c) LpTK 2 (SEQ ID NOS:10 and 20);
d) LpTK 3 (SEQ ID NO:12);
e) LpTK 4 (SEQ ID NO:14);
f) LpTK 13 (SEQ ID NO:16);
20 g) HpTK 5 (SEQ ID NO:24);
h) bpTK 1 (SEQ ID NO:25);
i) bpTK 2 (SEQ ID NO:26);
j) bpTK 3 (SEQ ID NO:27);
k) bpTK 4 (SEQ ID NO:28);
25 l) bpTK 5 (SEQ ID NO:29); and
m) bpTK 7 (SEQ ID NO:30).

6. A homogeneous protein of human origin which includes an amino acid sequence exhibiting sequence homology with the catalytic domain of tyrosine kinases of the c-kit family.
- 5 7. A homogeneous protein of Claim 6 in which the amino acid sequence is selected from the group consisting of:
 - a) SAL-S1 (SEQ ID NOS:6 and 18);
 - b) SAL-D4 (SEQ ID NO:8);
 - 10 c) LpTK 2 (SEQ ID NOS:10 and 20);
 - d) LpTK 3 (SEQ ID NO:12);
 - e) LpTK 4 (SEQ ID NO:14);
 - f) LpTK 13 (SEQ ID NO:16);
 - g) HpTK 5 (SEQ ID NO:24);
 - 15 h) bpTK 1 (SEQ ID NO:25);
 - i) bpTK 2 (SEQ ID NO:26);
 - j) bpTK 3 (SEQ ID NO:27);
 - k) bpTK 4 (SEQ ID NO:28);
 - l) bpTK 5 (SEQ ID NO:29); and
 - 20 m) bpTK 7 (SEQ ID NO:30).
8. A protein of human megakaryocytic origin which exhibits significant sequence homology with the FLT/FLK family of protein tyrosine kinases.
9. A protein of Claim 8 encoded by the nucleotide
25 sequence (SEQ ID NO:5).
10. A protein of Claim 8 encoded by the amino acid sequence (SEQ ID NO:6).

11. A protein of Claim 8 encoded by the nucleotide sequence (SEQ ID NO:17).
12. A protein of Claim 8 encoded by the amino acid sequence (SEQ ID NO: 18).
- 5 13. A protein of human megakaryocytic origin which exhibits significant sequence homology with the FGF receptor family of protein tyrosine kinases.
14. A protein of Claim 13 encoded by the nucleotide sequence (SEQ ID NO:7).
- 10 15. A protein of Claim 13 encoded by the amino acid sequence (SEQ ID NO:8).
- 15 16. A protein of human megakaryocytic or lymphocytic origin which exhibits significant sequence homology with the NGF receptor family of protein tyrosine kinases.
17. A protein of Claim 16 encoded by the nucleotide sequence (SEQ ID NO:11).
18. A protein of Claim 16 encoded by the amino acid sequence (SEQ ID NO:12).

19. A DNA expression vector containing a DNA sequence of human origin which hybridizes to a DNA fragment which hybridizes to DNA encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein kinases.
20. The DNA expression vector of Claim 19 containing a DNA sequence selected from the group consisting of:
- a) SAL-S1 (SEQ ID NOS:5 and 17);
 - b) SAL-D4 (SEQ ID NO:7);
 - 10 c) LpTK 2 (SEQ ID NOS:9 and 19);
 - d) LpTK 3 (SEQ ID NO:11);
 - e) LpTK 4 (SEQ ID NOS:13 and 21);
 - f) LpTK 13 (SEQ ID NO:15);
 - g) LpTK 25 (SEQ ID NO: 22); and
 - 15 h) HpTK 5 (SEQ ID NO: 23).
21. A cell transformed by the expression vector of Claim 19.

1/35

sals1 (160 bases)
FLKI-LIKE

1	PTKI/3 PRIMERS	21	41
s'ggatcctgtgcatcagtgacttagggctagggaacattctgtcgtcggaagcgatcgtggt			
		D P V H Q A L R A R N I L L S E S D V V	
61		81	101
gangatctgtgactttggccttgccctggcgacatctacaagaccccgactacgtccgcaa			
		K I C D F G L A: R D I Y K D P S (Y) V R K	
121		141	
gcatgcccggtgccctgaagtggatggcgccagaattc 3'			
		H A R L P L K W H A P E F	

FIGURE 1

sald4 (147 bases)-
FGFR-LIKE

1	PTKI/3 PRIMERS	21	41
5'	<u>ggatccattcacagagac</u>	<u>cttagcagcacgcaacatcctggctctcagaggacctggtaacc</u>	
	G S I H R D L A A R N I L V S E D L V T		
61		81	101
	<u>aaggtcagcgactttggcctggccaaagccgagcggaagggttagactcaagccggctg</u>		
	K V S D F G L A K A E R K G L D S S R L		
121	PTKKW PRIMER	141	
	<u>ccgtc aaatggatggctcccgaattc</u>		
	P V K W H A P E F		

FIGURE 2

3/35

LpTK2

GTTGGAATTCCTTCCGGCGCCATCCATTTACCGGCAGCTTTATTTTCGTGTCTAGATTCA
TAGATGTCTTCATTATCTACCTTAAAACTCTGGCAAGTCCAAAATCTGCTACTTTGTAG
ATATTATGTTACCAACGAGGACATTCCT

FIGURE 3A

LpTK3

GTGCACAGGGATCTCGCGGCTCGGAACATCCTCGTCGGGGAAAACACCCTCTCGAAAGTT
GGGGACTTCGGGTTAGCCAGGCTTATCAAGGAGGACGTCTACCTCTCCCATGACCACAAT
ATCCCCTACAAATGGATGGCCCCCTGAGGGAA

FIGURE 3B

LpTK4

GTTACCGAGATCTCAAGTCCAACAACATTTTGCTGCTGCAGCCCATGAGAGTGACGAC
ATGGAGCACAAGACCCTGAAGATCACCGACTTTGGCCTGGCCCGAGAGTGGCACAAAACC
ACACAAATGAGTGCCGC

FIGURE 3C

LpTK13

GTCAATCGTGACCTCGCCGCCCCGAAATGTGTTGCTAGTTACCCAACATTACGCCAAGATC
AGTGATTTCCGACTTTCCAAAGCACTGCGTGCTGATGAAAACTACTACAAGGCCAGACC
CATGGAAAAGTGGCCTGTCAAGTGGTACGCTCCGGAATGCATCAACTACTACAAGTTCTCC
AGCAAAAAGCGATGTCTGGTCCTTTGGAATTC

FIGURE 3D

1 TTCGACCTCG CCGACATTG ATTATTGACT AGTATTANT AGTATCAAT TACGGGGTCA TTAGTTTATA GCCATATAT GGACTTCCCG GTTACATAAC
AAGCTCCAGC GGGCTGTAACTAATACTGA TAATAACTGA TCAATAATTA TCNTTAGTTA ATGCCCCAGT ATCAAGTAT CCGGTATATA CCTCAAGGCG CANTGTATTG

101 TTACGGTAA TGGCCGCCCT GCGTGACCGC CCACGACCC CCGCCCATTG AGGTCAATA TACAGTATGT TCCATAGTA ACCCAATAG GGACTTTCCA
ATGGCATTT ACCGGGCGA CCGACTGGCG GGTGCTGGG GCGGGTAAC TGCAGTTATT ACTGCATACA AGGTATCAT TCGGTTATC CCTCAAGGCT

201 TTGACGTCAA TGGGTGGAGT ATTTACGGTA AACTGCCAC TTGGCAGTAC ATCAAGTGA TCATATGCCA AGTACGCCC CTATTACGT CAATCAGGT
AACTGCAGT ACCACACTCA TAAATGCCAT TTGACGGTG ACCCTCATG TACTTCACAT AGTATACGT TCATCGGCG GATAACTGCA GTTACTGCCA

301 AAATGGCCG CCTGCCATTA TGCCCACTAC ATGACCTAT GGCATTTC TACTGGCAG TACATCTAG TATTAGTAT CCTATTACC ATGGTATGC
TTACCGGCG GACCCGTANT ACGGTCTATG TACTGGATA CCTGAAAG ATGAACGTC ATGTACATG ATATCAGTA GGTATATGG TACCACCTAG

401 GGTTTTGGCA GTACATCAAT GGGCGTGGAT ACGGTTTCA CTCACGGGA TTTCAGTCTC TCCACCCAT TCGCTCAAT GCGATTGT TTGGCACCA
CCAAACCGT CATGTACTTA CCUGCACTA TCGCCAAT GAGTCCCTT AAGGTTTCA AGGTGGGTA ACCTCAGT CCTCAACA AACCTGTGT

501 AAATCAACGG GACTTTCCAA ATGTGCTAA CAATCCGCG CCAATGAGC TACGCTGTA CCGTGGAGG TCTATATAG CAGAGCTGT
TTACTTGGC CTCAAGGTT TTACAGCAT TTGAGGAGT GGTACTGCG GTTACCGCG ATCCGCAT CCCACCTCC AGATATATC GTCTCCAGCA

601 TTAGTGAAC GTACATCC CTTGACACG CATCCACCT GTTTGACCT CCATAGAGA CACCGGACC GATCCAGCT CCGCGGCGG GAACGGTGA
ATCAGCTTG CAGTCTAGC GACCTCTCG GTAGGTGGA CAAACTGA GGTATCTCT GTGGCCCTGG CTAGCTGGA GCGCGCGGCC CTTGCCAGCT

FIGURE 4A

5/35

701 TTGAAACGGG GATTCCCUGT GCCAAGAGTG ACGTAAGTAC CGCCTATAGA GTCTATAGGC CCALTGGCT TCGTTAGAAC GCGGCTACAA TTAAATACATA
AACCTTGGC CTAAGGGCA CGTTCTCAC TGCATTCTAC TGCATTCATG GCGATATCT CAGATATCCG GGTGAACCGA AGCAATCTTG CCGCGATGT ATTATGTAT

801 ACCTTATGTA TCATACACAT ACGATTAGG TGACACTATA GATAACATC CACTTGGCT TTCTCTCCAC AGGTGTCCAC TCCCAGGTCC AACTGCACCT
TGAATACAT AGTATGTGA TGCATAATCC ACTGTGATAT CTATTGTAG GTGAACCGA AAGAGAGGTG TCCACAGGTG AGGTCCAGG TTGACGTGA

901 CGGTCTATC GATTGAATTC CCGGGGATC CTCTAGAGAT CCTCGACCT CGAGATCCAT TGTGCTGGG CGATTCTTT ATCACTGATA AGTTGGTGA
GCCAAGATAG CTAACTTAAG GGGCCCTAG GAGATCTTA GGAGCTGGA CCTCTAGTA ACAGACCCG GCTTAGAAA TACTGACTAT TCAACCACT

1001 CATATTATGT TTATCAGTGA TAAAGTGTCA AGCATGACAA AGTTGCAGCC GAATACAGTG ATCGGTGCGG CCTAGACCT GTTGAACAG GTCCGCGTAG
GTATAATACA AATAGTCACT ATTTCACAGT TCGTACTGT TCAACGTGCG CTTATCTCAC TAGGCACGGC GGCATCTGA CAATCTGTC CAGCCGATC

1101 ACGTCTGAC GACAGGCMAA CTGGCGAAC GGTGGGGGT TCAGCAGCGG GCGCTTACT GGCACCTCAG GAACNACGG GCGTGTCTC AGGCACCTGC
TGCAGACTG CTGTGGGTT GACCCCTTG CCAACCCCA AGTGTGCGC CCGCAATGA CCGTGAAGTC CTGTGTGCC CCGACGAGC TCGTGACCT

1201 CGAAGCCATG CTGGCGGACA ATCATAGCAC TTGGGTGCG AGAGCCGAGG ACGACTGGG CTCATTCTG ACTGGGATG CCGCAGCTT CAGGCAGCG
GCTTCGGTAC GACCCCTCT TAGTATCTG AAGCCACGGC TCGCGCTGC TCGACCCG CAGTAAGAC TACCCCTAC GGCCTGCA GTCCGTGCG

1301 CTGCTCGCT ACCGCCAGCA CAATGGATCT CGAGGGATCT TCCATACCTA CCAATTCTGC GCGTGCAGGT CGCGCCGCA CTACTCTTG ATGTATTACT
GACGAGCGGA TGGCGTCTG GTTACCTAGA GCTCCCTAGA AGGTATGGAT GGTCAACAGC CGACGTCCA GCGCCGGCT GATGAGAAC TACATATCA

FIGURE 4B

1401 CATATTACCA AGGAATAACT GCGCGGCACA GGTCAGGTG CTGAAGGGAC ATTGTGAGAA GTGACCTAGA AGGCAAGAGG TGAGCCCTCT CTACCGCTG
GTATATGGT TCCTTATTGA CCGCCCGTCT CCAGTCCAC GACATCCCTG TACACTCTT CACTGGATCT TCCCTCTCC ACTCGGAGA CAGTCCGACC

1501 CATAGGGCC GCTTGGGGC TCTTTGCTCA AGCAGTACC CCAGTGTG GUAGGCACC TGTACTCAG CAGACCATGA AAGGCGTCT CCTTTCTCT
GTATCCCGG CGAATCCCG AGAATCCAGT TGTTCATTC GGTACAGAC CTTCCCTGG ACANTGACTC GTCTGTACT TTCCCGCAGA GCGAAGGAA

1601 GCAGCAGTCA GGGAACTC TGTCCACCA GCTTCTGTG GGAGCTTGA TATTATCAG GCGTCCCGG AGTCAATCCG AGGCTAACC CTTCCCTGT
CCTCTCAGT CCTTGTGAG ACCAGGTGT CCAAGAACAC CTTCCAGCT ATANTAGTC CGACCGGGG TCACTAGGC TCCGATTGG GGAGGACAC

1701 GTGCTCAGT GGTACACTC CTTGCTCCTT TCAATGCTCC TCTTGGCTC CTGTTCTC TTGAAGTT GTAGTAGATA GCAGAGAAA TAGCAATGT
CAGCAACTCA CCAGTGTGAG GAACAGGTGA AGTACGAGG AGAACGGAG GACCAAGGAG AACCTTCAAA CATATCTAT CGTCTCTT ATCGCTTCA

1801 CTTAAGTCT TTGATCTTC TTATAGTGC ACAGAGAAA TGTGACGTA TGTGCTTC TCTCTCTG CTACAGTAC CTGAAGCCG TTCTTGTCT
GAATTTTCA AACTAGAAAG ATATTACG TCTCTCTT ACCACTGCAT ACGACGGAG ACAGAGAGC GAATCGATG GACTCGGG AAGAACAGA
O R P G S E Q R
349

1901 ATACCTGCTC TCTATCTGT CACATCTCTC CGAGGCCAGC ACGTCCAC TGTCTCTG GTTCTCACA GAGCTTGT AGTCTGTGG GTCATGGG
TATGGACGAG AGATAGACGA GTGTGAGGAG GCTCCGCTG TGTAGGCTG ACAGACAGC CACAGGTGT CTCCGAAACA TCCAGCAACC CCACTACCC
341 Y R S E I Q E C E E S A L V H C S D T Q N D V S G K Y T T P T H P

2001 AATTCCTCAA ATGTCTTCT CTTGGAGGAA CCACGGGTCT CAGCCCTCT GCGCAGGCAC CCGGGAAGG ACACCACTT GTATACCTG GCGCCAGG
TTAAGGAGTT TACAGAATA GCACCTCTT GTGCCCCA GTCGGGAGA CCGTCCGTG GCGCTTTCC TGTGGTCAA CATATGGAC GCGCGTCCG
308 P E F F T K H R S S C R T E A G R A L C G P F S V W N Y Y R A A L

FIGURE 4C

2101 TGTGGCGGTG CAGGTTGGC GGGTGTCTT CAGCGTCAGC TGTGGGCACT TGTAGGGCCA TGTGGCACG CTGGCAGAG CTGCGCTCTT CTGAGCTCTG
 ACACCGCGAC GTCCGACCG CCGACAGCA GTCCGACGTG GACCGGTAC ACATCCGCT ACCACTGTG GACGCTTC GACGGAGAA GACTCGAGAC
 215 S H R Q L S P P S D E A D A Q A I H L A M T S V Q S F S G E S S O

2201 AGAGCTGGC GGGGCCATGC AATCTCTC TTCTCTTG AGCCCTGC COTGGAGCAG GTCCCCAGG ATCTCCACCA GTTCGAGAA TGCAGGTCTC
 TCTCGACCG CCGCGTACG TCTGGAGAG AAGGAGAAC TCCGGGACG GACCTCTC CAGCGTCC TAGAGTGGT CGAGCTCTT AGTCCAGAG
 225 S S R P A M C V E E E Q L G R G Q L L D G L I E V L E S F A P R

2301 CCCTTGGGT CTCCGGACCA GCACTTCCG AATCTTCCG GTATGGGCGG ATGGGCCAGC TCCGGGCGG TCATCTTGT CCGTCTCTC AGCGCTGGC
 CGAACCCCA GAGCTGTG GTCAAGTCG TACTACCGG CATACCGCC TCACCGGTG AGGCTCCGG AGTAGNACA CGGCAGAG TCGCGACCG
 235 A K P D G S W C N L M I R R I A P T A L E P A R H R T G D R L R Q

2401 AGAATCTCTC ATGTATCTG ACCCAGGCT ACAGGAGGC CCGCAGAG AGATCTTCC AGAGAGAGC CCGAAGGAC CACAGCTCAC TCTCGTGT
 TCTGAGGAG TACTAGACG TGGGTCCCA TCGCCCTCG GGGTCTCTC TCTAGAGG TCTCTCTG GGTTCCTG GTGTCAGTG AGACGACCA
 245 C F E E N I Q V G P Y P S A G L S P I E W L L V G P S W V D S O T T

2501 GTACACUUG TCGAAGTGC TTTCAGGGC CATCCTCTC AGGGCAGCC GGCCTTGC GGCCTGCG GACCTCTG ATCAGCCCA GAAATCTA CAGGCUUGT
 CATGTGAGC AGCTCTACG AATCTCCG GTAGTCAAG TCCCGTCCG CCGTGACG GACCTCTG GACCTCTG ATCAGCCCA GAAATCTA CAGGCUUGT
 255 Y V K D F Y S E P A H W K L P L N A S G K R V Y D P D K Y I D R A

2601 AGGCCAATCT CACAGTCTT CACAGCTG CTCTCCGACA GCAGATCTT CCGACAGCC AGGTCTCTG GATGCACT TCGGAGGCC AGAATCTCA
 TCCGTTTCA GTCTACGA GTCTGACG GAAAGCTGT GTCTTACA GGTCTGCG TCCAGAGCA CCTACGTGA AGCCCTCG TCTTGAGGT
 265 L G F D C I K V V D S E S L L I M R A A L D R H I C K R S A L P E

2701 TCCCTCTG CACCTGGAG CTGTACGAGA CAAGATCTC CATGTGAG GGGTCAGC ACAGTCTC AGTCTCTG TCTGGAGAG CCGCTCTG
 AGGAGAGCG GTGACCTTC GACATCTCT GTTCTAGG GTACAGTC CCGAGTCG TCTCCAGCAG TCGAAGAC AGACTCTC GGGCGAGCG
 275 M G R A V O P S Y C V L D E H T L P S L W L D E A E Q D P S A R R A

FIGURE 4D

FIGURE 4E

2801 TCCGCCCTCG CTCTTCAGAA ACCGGGCGAA GAGGACCTTG TCGTGTCTCC CCGCGCGGCT CCGATCCAGC CTGGCCAGCT CCACATGGC GCGAAGCGT
 AGCGGGAGC CAGAGCTCT TGGCGCGCTT CTCCTGGAC AGCGACGAGG GCGCGGCGGA GGTAGGTG GACCGCTGA GTGTACCG CGCTTCGA
 41 C G E T K S F R A F L V R D S S G P R R D L R A L E V M A R F R

2901 CCGCGTCTCT CCGCAGACTT CTCTGCGGA TGCACGAAGC TGGCTCGAGG GCGCCAGTC GTCCGCCGA GAGGCGCTC CATTCGCCG CCGCCCGCG
 GCGCGACCA GCCCTCTGA GAGGAGGCTT AGTGTCTG ACCGAGCTCC CCGCGTCTAG CAGCGCGCT CTCCGGGAG GTAGGGGC GCGCGCGCC
 8 C R Q E P S K E

3001 CCGCGCGCAG CCGCGCGCT CACCGGCGAG GCGTGGCG CCGACTCTA GAGTGCACCT GCAGAGCTT GCGCGGCTG GCGCAGCTG TTTATTCGAG
 GCGCGCGCT CCGCGCGGA GTGGCTCT CCGGAGCGG GCGTGCAGT CTCAGCTGA GGTCTCGAA CCGCGGCTAC CCGGTGAA CCGGTGAA

3101 CTTATTAATG TTACAATAA ACCAATAGCA TCACAATTT CACAATAA GCATTTTTT CACTGCATC TAGTGTCTT TTGTCCAAC TCATCAATCT
 GAATATTACC ATGTTTAT TCGTATCTT AGTGTATAA GTGTTATTT CGTAAJANA GTACGTAAAG ATCAACACCA ACAGGTTTG AGTAGTTACA

3201 ATCTTATCAT GTCTGGATCG ATCGGGAATT ATTCGGCGC AGCACCATGG CCGCAATAA CCGTGAAG AGGAATTTG TTAGTACCT TCTGAGCGG
 TAAATAGTA CACACCTAGC TAGCCCTTAA TTAAGCGCG TCGTGTACC GCACTTTAT GGACACTTC TCCTGAAAC ATCCATGGA AGACTCGCC

3301 AAAGAACCAG CTGTGAATG TGTGTACGTT AGGTGTGGA AAGTCCCGC GCTCCCGAG AGCAGAAAT ATCAAGCA TGCATCTCA TTAGTCAGCA
 TTTCTTGGT CACACCTTAC ACACACTCA TCCACACCT TTCAGGGGTC CGAGGGGTC TCGTCTTCA TAGCTTCTT ACGTAGATT ATCAGTCTT

3401 ACCAGGTGTC GAAGTCCCG AGGTCCCGA CAGGCGAGAA GTATGCAJAG CATGCATCTC AATTAGTCAG CAACCATAGT CCGCGCGCTA ACTCGCGCCA
 TGTCCACAC CTTTCAGGGG TCCGAGGGT CCGCGCTT CATACGTTTC GTACGTAGG TTAATCAGTC GTTGTATCA GCGCGCGAT TCAGGCGGT

9/35

FIGURE 4F

3501 TCCGCCCCCT AACTCCGCC AGTTCGCC ATTCGCC CCATGCTGA CTAATTTT TTAATTATGC AGAGGCCGAG GCGGCTCGG CCTCTGAGCT
AGGGCGGGA TCGAGGCGG TCAAGGCGG TAAGAGCGG GUTACCGACT GATTAAAAA AATAAATACG TCCTCGGCTC CCGCGAGCC GGAGACTGA

3601 ATTCCAGNAG TAGTGAGGAG GCTTTTGG AGGCTAGGC TTTTGCAAAA AGCTGTAAAC AGCTTGGCAC TGGCGTCTGT TTACAAACGT CGTGACTGGG
TAAGGTCTTC ATCACTCTCTC CGAAAAACC TCCGATCCG AAACGTTTT TCGACAAATG TCGAACCGTG ACCGACCA AATGTGCA GCATGACCC

3701¹ AAACCCCTGG CGTTACCCAA CTTAATCCG TTGACGACA TCCCCCTTC GCCACTGGC GTAAAGCCA AGAGCCCGC ACCATCGCC CTTCCCAACA
TTTTGGACC GCANTGGGT GAATTAGCG AACGTCTGT AGGGGGAG CCGTCGACCG CATTATCGT TCCTCGGCG TGCTAGCG GAAGGTGT

3801 GTTCGGTAGC CTGANTGGG ANTGGCCCT GATGGGTAT TTCTCTTA CGCATCTGT CGGTATTCA CACCGATAC GTCAAGCAA CCATAGTACG
CAACGCATCG GACTTACCG TTACCGCGA CTAGGCCATA AAAGAGGAT GCCTAGACAC GCCATAAGT GTGGGTATG CAGTTCTGT GGTATCATGC

3901¹ CGCCTGTAG CGGCGATTA AGCGGCGG GTGTGGTGT TAGCGGCGA GTGACCGCTA CACTGCGCA GCGCCTAGCG CCGCTCTT TCGCTTCTT
CGGGACATC GCGCGTAA TCGCGCGCC CACACCACTA ATCGCGCTG CACTGGAT GTGACGGTG CCGGATCG GCGGAGAA ACGAAGAA

4001 CCTTCCTTT CTGCGCACGT TCGCGGCT TCCCGTCAA GCTCAATC GGGGCTCC TTATGGTTC CGATTATG CTTACGCA CTTGACCCG
CGGAAGGAAA GAGCGGTGA ACGGCGGA AGGGCAGT CGAGATTTAG CCCCAGGG AATCCCAAG GTAAATCAC GAATGCCOT GGAGCTGGG

4101 AAAAAGCTT ATTGGGTGA TGCTTACGT AGTGGGCAAT CGGCTGATA GAGGTTTTT CCGCTTTGA CGTTGGACTC CAGCTCTT ATAGTGGAC
TTTTTTGAC TAACCCACT ACCAAGTGA TCACCGGTA CGGGGAACT CGAACCTCAG GTCCAGAAA TTATCACCTC

10/35

FIGURE 4G

4201 TCTGTGTCCT AACTGGACCA ACACTCAACC CTATCTGGG CTATCTTTT GATTATTAAG GATTTTGGC GATTTCGGC TATGGTAA AAAATGAGCT
ACAACAAGGT TTGACCTTGT TGTGAGTGG GATAGAGCC GATAAGAAA CTAATATTC CTAAGCCCG ATACCAATT TTTTACTCGA

4301 GATTTAACAA AATTAAACG CGAATTTAA CAAATATTA ACGTTTACAA TTTATCGTG CACTCTCAGT ACATCTGCT CTGATGCCCG ATAGTTAAGC
CTAATTTGTT TTTAANTGC GCTAANAAT GTTTTATAAT TGCAATGTT AANATCCAC GTGAGAGTCA TGTAGACCA GACTACGGC TATCAATTCC

4401 CAACTCCCT ATCGCTACGT GACTGGGTCA TGGTGGCC CCGACACCG CCAACACCG CTGACGGCG GACTGCCCG TGTCTGCTCC CGGATCCCG
GTTGAGGCG TAGCGATCCA CTGACCCACT ACCGACCGG GCTGTGGC GCTGTGGC GACTGCCCG GACTGCCCG ACAGACGAG GCGTAGGCG

4501 TTACAGACAA GCTGACCG TCTCCGGAG CTGCATGCT CAGAGGTTT CACCTCATC ACCGAACG CCGAGGAGT ATCTTCAAG ACCAAGGCG
AATCTCTTT CGACACTGGC AGAGGCCCTC GAGTACACA GTCTCAAAA GTGCGTAG TGGTTTGG CCGTCCGTC TAAGAACTTC TGCCTTCCCG

4601 CTCGTGATAC GCTATTTT ATAGCTTAAT GTCATGATAA TAATGGTTT TTACAGCTCA GTTGGCACT TTGGGGGAAA TGTGGCGGA ACCCTATT
GAGCACTATG CGGATAAAA TATCCANTTA CAGTACTATT ATTACCAAG AATCTCAGT CCACCGTGA AGCCCCCTT ACACCGGCT TGGGNTAA

4701 GTTATTTT CTAATACAT TCAATATGT ATCGGCTCAT GAGACATAA CCTGATAA TGCTTCAATA ATATTGAAA AGGAAGAGT TGAATATTCA
CAATATAAA GATTATGTA AGTTATACA TAGGCGATG CTCTGTTAT GCGACTATT ACCAAGTAT TATACTTTT TCCCTCTCAT ACTCATAGT

4801 ACATTTCCGT GTCGCCCTTA TCCCTTTT TCGGGCATTT TCGCTTCTC TTTTGTCTA CCCAGAACG CTGGTCAAG TAAAGATGC TGAAGATCAG
TGTAAGGCA CAGCGGGAAT AGGGAAAA AGCCGTA ACGGAGGAC AAAAAAGT GGTCTTTC GACCCTTTC ATTTCTACG ACTTCTAGT

11/35

FIGURE 4II

401 TTGGGTCAC GACTGGGTTA CATCGAATC GATCTCAACA GCGTAAGAT CTTGAGAGT TTTCGCCCG AGAACGTTT TCCANTGATG AGCACTTTT
AACCCACGTG CTCACCCCAAT GTAGCTTCAC CTAGAGTTGT CCCCATTCTA GGAATCTCA AAAGCGGCG TTCTTGAAA AGGTIACATC TCGTGAAA

501 AACTTCTCT ATGTGGCGCG GTATTATCC GTGATGACG CACTACTGCG CCGGCAAGAG CAATCGGTC GCGGCATACA CTATTCTAG ATGACTTGG TTGAGTACTC
TTCAAGACGA TACACCGCG CATATAGG CACTACTGCG CCGGTTCTC GTTAGCCAG CGCGTATGT GATAGAGTC TTACTGAACC AACTCATGAG

5101 ACCAGTCACA GAAAGCATC TTACGGTGG CATGACAGTA AGAGATTAT CCAGTCTGCG CATACCATG AGTATAACA CTCGGGCCAA CTIACCTCTG
TGTCACTGT CTTTCGTAG ATGCTTACC GTACTGTCT TCTTTAATA GTCACGAG GTATTGGTAC TCACTATTGT GACCCCGGT GAATGAAGAC

5201 ACAAGCATCG GAGGACCGAA GGAGCTAACC GCTTTTTCG ACACATGGG GGATCATGTA ACTGCTTG ATCGTTGGG ACGGAGCTG AATGAAGCA
TGTCTAGC CTCCTGGCT CCTCGATTGG CGAAAAACG TGTGTACCC CTAATACAT TGAGCGAAC TAGCAACCT TGGCTCGAC TTACTTCTG

5301 TACCAACGA CGAGCTGAC ACCAGCATG CAGCAGCAAT GGCAACAAC TTGGCAAC TATTACCTG CGAATCTCT ACTCTAGCT CCGCGCAACA
ATGGTTTGT GCTCGACTG TGGTGCTAG GTGCTGTTA CCGTGTCTG AACGCTTG AATATGACC GCTTATGAA TCAGATGAA GCGCCGTGT

5401 ATTAATAGAC TGGATGGAG CGGATAAGT TGCAGGACCA CTCTGGGCT CGGCTTTC GCGTCTGCG TTATTCTG ATAAATCTG ACCGGTCTG
TAATTATCTG ACCTACCTCC GCTATTCTA ACCTCTCT GAAGACGGA GCGGGAAGG CCGACCGACC AAATAACGAC TATTAGACC TCGGCCACTC

5501 CGTGGTCTC GCGTATCAT TGCAGCACTG GGGCCAGATG GTAGCCCTC CCGTATGTA GTATCTACA CGACGGGAG TCAGCCAACT ATGCATGAAC
GCACCCAGAG CCCCATAGTA AGCTGCTAC CCGGCTCTAC CATTCGGAG GGCATAGCT CATTGCTC AGTCCCTC TACCTACTC

FIGURE 4I

5601 GAAATAGACA GATCGCTGAG ATAGGTGGCT CACTGATTA CCATTCTTAA CTGTCAGACC AACTTTACTC ATATATACTT TAGATTGATT TAAACTTC
CTTTATCTGT CTACCGACTC TATCCACGGA GTGACTAATT CGTAACCAAT CACAGTCTGG TTCAANTCAG TATATATGAA ATCTAACTAA ATTITGAK

5701 TTTTAAATTT AAAGGATCT AGGTGAAGAT CCTTTTIGAT AATCTCATGA CCAAAATCCC TTAACGTGAG TTTTCGTTC ACTGAGCCTC AGACCCCGTA
AAAAATTAAA TTTTCTAGA TCCACTTCTA GGAAAACTA TTAGAGTACT GCTTTTAGGG AATTGCACCTC AAAAGCAAGG TCACTCGCAG TCTGGGRCAT

5801 GAAAAGATCA AAGGATCTTC TTGAGATCTT TTTTCTGC GGTAAATCTG CTGCTTGCAA ACAAAAAC CACCTCTACC ACCGTGGTT TGTTCGCGG
CTTTCTAGT TTCTAGAG AACTCTAGGA AAAAAGACG CGCATTAGAC GACGAACGTT TGTTTTTTIG GTCCCATGG TCGCCACCAA ACNAACGGCC

5901 ATCAAGAGCT ACCAACTCTT TTTCCGAAGG TAACTGGCTT CAGCAGAGCG CAGATACCAG ATACTGTCTT TCTAGTCTAG CCGTAGTTAG GCCACCCTT
TAGTCTCGA TGGTTGAGAA AAGGCTTCC ATTGACCGAA GTCTCTCGG GTCTATGGTT TATGACAGGA AGATCAGATC GCATCAATC CGGTGCTGAA

6001 CAAGAACTCT GTACACCGC CTACATACCT CGCTCTGCTA ATCTGTATG CAGTGGCTGC TCCACGTGGC GATAAGTGT GTCTTACCG GTTGGACTCA
GTCTTGAGA CATCTGCGG GATCTATGGA GCGAGACGAT TAGACAAATG GTACCGAGC ACGGTCAACG CTATTACGA CAGATGGCC CAACCTGAGT

6101 AGACGATAGT TACCGATTA GCGCGACGG TCGGGCTGAA CCGGGGGTTC GTGCACACAG CCCACCTTGG ACGCAACGAC CTACACCGAA CTGAGATACC
TCTGCTATCA ATGGCTATT CCGGTGCGC AGCCCGACTT GCGCCCAAG CACGTGTGC CGGTGGAAC TCGTTCCTG GATGTGGCTT GACTCTATGG

6201 TACAGCGTGA GCATTGAGAA AGCGCAGCG TTCCCGAAGG GAGAAAGGG GACAGCTATC CCGTAAGCGG CAGGGTGGGA ACAGGAGAGC GCACGAGGG
ATGTCGCAC CTGAACCTCT TCGCGGTGCG AAGGGCTTCC CTCTTCCG CTGTCCATAG GCCATTGCGG GTCCAGGCT TGTCTCTGCG CGTCTCTCC

FIGURE 4J

6301 GCTTCCAGGG GGAAAGCCTT GGTATCTTAA TAGTCTGTG GGGTTTCGCC ACCCTGACT TGAGGTGCGA TTTTGTGAT GCTGCTCAGG GGGCGCGAGC
CGAAGTCCC CCTTTGCGGA CCATAGAAAT ATCAGGACAG CCCAAAGCG TGGAGACTGA ACTCGAGCT AAAACACTA CGAGCAGTCC CCGCGCTCG

6401 CTATCGAAA AGCCAGCAA CCGGGCTTT TTACGTTCC TGGCTTTG CTGGCTTTT GCTCAGATGT TCTTCTCTC GTTATCCCT GATCTCTCG
GATACCTTT TCGGTCTGT GCGCGGAAA ATGCCAAGG ACCGAAAC GACCGAAA CGAGTGACA AGAAGGAGG CAATAGGGA CTAAACACC

6501 ATACCCAT TACCGCTTT CAGTGACTG ATACCGCTG CCGCAGCGA ACCACCGAG CGAGCGATC AGTGACGAG GAACCGAG AGCCCCCAAT
TATTGGCATA ATGGCGAAA CTCACTCGAC TATGGCAGC GCGTGGCT TGTGGCTG CGTGGCTC TCCTGCTC TCGGGGTTA

6601 ACCCAACCG CTTCTCCCG CCGTTGGGC GATTCTTAA TCCAGCTGGC AGCAGAGTT TCCCGACTGG AAGCGGCA GTGAGCGAA CGCAATTAAT
TGCCTTGGC GGACAGGGC GGCACCGG CTAAGTAAT AGCTCGACG TGTGTCCA AGGCTGACC TTTCGCCCT CACTCGCTT CGTTTAATTA

6701 GTAGTTACC TCACTATTA GGCACCCAG GCTTACACT TTATGCTTC GCTCGTATG TTGTGTGAA TTGTAGCGG ATACATTT CACACAGAA
CACTCANTG AGTGAGTAAT CCGTGGGTC CGAATGTGA ATACGGAAG CCGAGCATAC AACACCTT AACACTGCC TATTGTAAA GTGTGCTT

6801 ACAGCTATG CCATGATTAC GAATTA
TCTGATACT GGTACTATG CTTAAT

14/35

FIGURE 5A

1 TTGAGCTCG CCGACATG ATATTTGACT AGTATATTAAT TACGAGCTCA TTACTTTCATA GCCCATATAT GGACTTCCGC GTTACATATAC
AAGTTCAGC GGGCTCTAAC TAATAACTGA TCATAATATA TATGCCCCAGT ATTCACAGTAT CCGGTATATA CCTCAAGGCG CAATGTATTG

101 TTACGTTAA TGGCCCGCT GGTACCGC CCACGACCC CCGCCCATTTG ACCTCAATAA TCACGTATGT TCCCATAGTA ACGCCATAG GGACTTTCCA
AATGCCATT AT ACCGCGCGA CCGACTGGG GCTTCTTGG GCGGGTAAC TCACGTTAT ACTGCATACA AGGTATCAT TCGGTATAT CCTGMAAGGT

201 TTGAGCTCA TGGTGGAGT ATTTACGTA AACTGCGAC TTGCGAGTAC ATCAAGTGA TCATATGCCA AGTACGCCC CTATTACAT CAATGACGGT
AACTGAGTT ACCCACCCTA TAATGCCAT TTGACGGTG AACCTCATG TAGTTCACAT AGTATACGGT TCATGCGGG GATACTGCA GTTACTGCCA

301 AATGSCCG CCTGGCATT TGGCCAGTAC ATGACCTTAT GGGACTTTCC TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGCTATGC
TTTACCGGC GCACCGTAAT ACGGTCTATG TACTGGAATA CCTGMAAGG ATGAACCGTC ATGTAGATGC ATAATCAGTA GCGATAATYG TACCACCTAGC

401 GCTTTTGGCA GTACATCAAT GGGCTGGAT AGCGTTTGA CTCACGGGA TTTCCAAGTC TCCACCCCAT TGACGTCAAT GGGAGTTTGT TTGCGACCA
CCAAAACCGT CATGTAGTTA CCGCACCTA TCGCCAACT GAGTCCCCCT AAGGTTTCA AGGTGGGTA ACTGCAGTTA CCTCAACA AAACGTTGT

501 AATCAACGG GACTTTCCAA AATGCTGTA CACTCGCC CCATTACGC AANTGGCGG TAGCGGTGA CCGTGGAGG TCTATATAG CAGAGCTCGT
TTTACTTGG CTGAAGGTT TTACAGCAT GTTACGCGG GGTACCTGG TTTACCGCC ATCCGCACAT GCCACCTCC AGATATATTG GTCTCGAGCA

601 TTAGTGACC GTCAGATCG CTGGAGCG CATCCACCT GTTTGACCT CCATAGAAGA CACCGGAGC GATCCAGCCT CCGCGCGCGG GAACGTTGCA
AATCACTTG CAGTCTAGG GACCTCTGG GTAGGTGCA CAAMCTGGA GGTATCTTCT GTGCCCCCTG CTAGGTGGA GCGCGCGCC CTTCACCGT

701 TTGAGCGCG GATTCGCCG GCCAAGAGT ACCTAGTAC CGCTATAGA GTCTATAGC CCACCTGGT TCGTTAGAAC GCGGTACAA TTAATACATA
AACCTTCCG CTAAAGGCA CGTCTCTAC TGCATTCATG GCGATATCT CAGATATCCG GGTGACCGA AGCAATCTTG CCGCGATGTT AATATGAT

15/35

FIGURE 5B

801 ACCTTATCTA TCATACACAT ACCATTTAGG TGACACTATA GAATACATC CACTTTCCCT TTCTCTCCAC AGGTGTCAC TCCCAGGTCC AACTGCACCT
TGAATATACAT AGTATGTCTA TGTAAATTC ACTGTGATAT CTATTGTAG GTGAAACGGA AAGAGAGGTG TCCACAGGTG AGGTCCAGG TTGACGTGGA

901 CGGTTCTATC GATTGAATTC CCGGGGATC CTCTAGAGAT CCTCGACCT CGAGTCGACT TTTTCTTTT TTTTGTAGG CCAAGGGTA CTCTTTTTT
TCCAGATAG CTAACTTAAG GGGGCTTAG GAGATCTCTA GCGAGCTGA GCTCAGCTGA AAAAAAANA AAAAAATCC GGTTCCTCAT GAAGAAAAAG

1001 TTTTATTAAT ACTCAGAGT CTAGGCCACA GCAATCTACT GTTCTCTCT CATTTCCTTA AACTATTTG ATACCTATTT CTCAGACTTT ATGGGCTATT
AAATAATTA TGAATCTTCA GATCCGGTGT CGTTAGATGA CAAGAGAGA GTAAAGGAT TTGATAAAG TATGGATANA GAGTCTGANA TACCCGATNA

1101 AGACATTCT CACATTTCCTA TAGATAATM CTCATCCGT TTGCAACCTG ATTCTCAATA TTAAGAGATT AAACTAATG TATAGACTC TCAGTTGACA
TCTCTAAGA GTGTAAAGT ATCTATTATT GAGTACCTA AACGTTGAC TAAGAGTTAT AATCTCTAA TTTTCTTAC ATATACTGAG AGTCACCTGT

1201 CATACGAG TACAGAAAAA TTCCATCAT TTCTCTGCA AATGAAAAA GACTTCGTT TCTCAACAGC TGCATCATTT TTTTATGAT AGAAAAAAT
CTATGACTTC ATGTCTTTT AAGGTAGTAA AGGAAGAGCT TTACTTTTT CTGAAGCANA AGAGTTCTCG ACCTAGTANA AAAATACGTA TCTTTTTTA

1301 GTGCAATTAC TCCAAGTACA ATCAAGTCAT TTACATGGC TTACCATCA TTCTAGTTAC AGGATATTTT AAAAGAGAAA AAAAATCTC AAAGCACAGG
CAGCTTAATG AGGTTTCATGT TAGTTTCAGTA AATGTACCG AATGGTAGT AACATCAATG TCCATANA TTTTCTCTT TTTTGTAGG TTTCTGTCC

1401 TCCCTGCTGT CAGCAAGCA ATCAAAATCC TTCTAATMA CAGCTGATG GGATTCAGCA ATCTGAGGAA TAATGATAA CCACTCTAAT CAGTAAACAG
AGACGACAC GTCGTTTCGT TAGTTTAAGG AAGTATTATT GTGGGACTAC CCTAAGTCTG TAGACTCTT ATTACTTATT GGTGAGATTA GTCATTTGTC

1501 GAAATGCTA CAACAGTCAG TGAATAAANA TTGACTATC ATCTGTGAT TCTCTTCTG GACATTTCAA ACAATAAATG GAATGTAG TATCTCTTAA
CTTTTACAT GTGTCTCATG ACTCTATTTT AACCTGATG TAGACMACTA AGAGAACTAG CTGTAAAGTT TGTATTATAC CTTTACATTC ATAGACATTT

16/35

FIGURE 5C

1601 AAAGAAAAT AACTGGTTT AGTGTGCTTA ATTATTACAG GCAGTGAGGA AATTATATAT CACCTTCACT GTCTGCAGT GTTCCCACT CAATAAAATG
TTCTTTTITA TTGAACCAA TCACACGAAT TAAATATGTC CGTCACTCTT TTAATATATA GTGAACTGA CAGGACGTCA CAACGGGTCA GTTATTTTAC

1701 CACAAAATAT CTTTTCATA ATACATGACC AACTTTATCC TATCACTTGA ATATGTCAGG ATAAACTGAT TGTGCAGTTG GTTGATAACA TTGTATTTTG
CTCTTTTATTA GAAAAGTAT TATGTACCGG TTGAANTAG ATAGTAACT TATACAGTCC TATTTCAGTA ACACGTCAAC CAACTATTGT AACATNAAC

1801 GAATGGATTA TTGTGAATTTG TTTTGTACT TTTTATTTG ATATTCTCT CCAGTGTCA TCTTATGAG TTATTTCAT CTGAATATGA AGAGTCTGTT
CTTACCTAAT AAACTTAAAC AAACGATGA AATAATAAAC TATNAGAGA GGTCACAAGT AGAATACTTC AATAAACGTA GACTTATCT TCTCAGACAA
506 O R I P N N A D S Y S S D T

1901 TCAAAATAGT CTTCAAGTTT CCAACGCAGT GTCTCAATG TAGTGTCTT CTAGGCTCT GCATCCAGC ACTCCAACAT GATGTGTGTA AATTGCTGTG
AGTTTATCA GAAGTTCNAA GGTTCGTCA CAGAGTTTAC ATCCAGCAG GAATCCGAGA CGTAAGGTCTG TGAGGTGCTA CTACAACATT TTAACGACAC
492 E F Y D E L K W R L T E F T P R E A N W C E L H I N Y F Q Q

2001 GACAGTTGGA TGCTTGCGGA AGTCTATAGT TTTGAGCCAA CATCTGANTT ACCTGGGCAC CTCTCATACC ACTGTNAGGC ATTTTGGCAT AAGTAATGAT
CTGTCAACCT ACCAACGCCT TCAGATATCA AAACCTGGTT GTAGACCTAA TGGACCCCTG GACAGTATGG TGACATTCCG TAAACGGTA TTCATTTACTA
459 P C N S P Q P L R Y N Q A L M Q I V Q A G T H G S Y P M K G Y T I I

2101 TTCATNAGA AGGATTCCAA ATGACCATAC ATCGGACTTA ATGCTGAATT TATTACTAGC ATGCGCTTCG GCGCAGTCC ACTTCACCGC CAGCTTTTATT
AAGTATTTCT TCCTAAGCTT TACTGTATG TAGCCTGAAT TACGACTTAA ATAAATGATGC TTACCGAAGC CCGGTCAAG TGAAGTGGC GTCGAANTAA
425 E Y L L I G F S W V D S K I S F K N S R I A E P A T W K V P L K I

2201 TCGTGCTAG ATTCAATAGT GCTTTCATTA TCTTACCTTAA AAACCTGTCG AAGTCCAAA TCTGTACTTT TGTAGATATT ATGTTACCA ACGAGGACAT
AGCACAGATC TAAGTATCTA CAGAATMAT AGATGGAATT TTTGAGACCG TTCAGCTTTT AGACGATGA ACATCTATAA TACAAGTCT TCGTCTCTGA
392 E H R S E Y I D E N D V K F V R A L G F D A V K Y I N H E G V L V

2301 TTCTGACAGC CAGATCTCTG TGAATGTAGT TCCGAGACTC CAGATAGGCC ATTCCAGAGG CAACCTGTGC CGCCATGTCT ACCTGTTGAG TCAGATTCAT
AAGACGCTCG GTCTAGAGAC ACTTACATCA AGCTCTTAG GTCTATCCCG TAAAGTCTCC GTTGCACACG GCGGTACAGA TGGACAACTC AGTCTACCTA
359 N R A A L D R H I Y N R S E L Y A H G S A V Q A A H D V Q Q T L H I

17/35

FIGURE 5D

2401 TTCTGATCCA GGTCAATTTT GGAGATATTC TTGACAGCTT CCATGCTCA TCAACTCTGT AATAATATAA ATTGATCTT CTAAGTCCA AACACATAA
 AMAACTAGCT CACAGTAAA CCTCTATAAG AACCTGAA GGTACAGAGT AGTTACAGCA TTATTATATT TAACCTAGAA GATTTCAGCT TTCTGTATT
 325 K S G T D N Q L Y E Q L S G H R M L E T I I Y I P D E L T C V A Y

2501 AGCTGGATAA GCTTTGGATG TCTTAGGTTT TCAATTATCT GTGCTCTCCT CAGGAAGTCA TTGGAATCCA TTGAACCTGG TTTTAATGTT TTCACTGCTA
 TCGACCTATT CGAACCTAC AGAATCCNAG AGTAATAGA CACGGAGGGA GTCTTTCAGT AACCTAGGT AAATTTACAA AAGTGACGAT
 292 L Q I L K P H R L N K M I Q A E R L F D N P D H S G P K L T K V A

2601 CTGGAGTGGT ATTGTTCCAC AGACCTTCCC ATACTTCCC AACTGACCA GATCCCAATC GCTTCAGNAG CTGTATGGAG TTGCGGTCTA TCTCCCATTTG
 GACCTCACCA TAACAAGGTG TCTGGAAAGG TAAGAAGCGG TTGACTGGT CTAGGGTTAG CGAAGTCTTC GACATACCTC AAGCCAGAT AGAGGTAAAC
 259 V P T T N N W L G E W V E G F Q G S G L R K L L Q I S N R D I E W Q

2701 GTCCACGGTT TTATACGACA ATCAAAATGG AGCTGACCC TGGATCTTTA AGCATGGTTT CCCAGCTTTG ACACACAGGC CGTCACTTGT CTTCGTGTAG
 CAGGTGCCAA ATATCTCTGT TTAGTTTACC TCGACCTTGG ACCTAGAAAT TCGTACCANA GGGTCCGAC TGTGTGTCCG GCAGTGAACA GAACACATC
 225 D V T K Y S L D F P A P V Q I K L C P K G L K V C L G D S T K T Y

2801 TGCTCACAA ATTGCTTCAG TGTGAAAG ATTCTTCTTC GCGTGAAGA AAATCCCTT TCAATCCAGTC TTTTAATTTT GTAGTGTTTT ACAACTGCTC
 ACCGAGTGT TAAGCAAGTC ACAACTTTT TAAGAAAGAG CGCACTTTT TTAGCGGA AGTAGGTCAG AAATTTAGA CATCACAAA TGTTCACGAG
 192 H S V F E N L T S F I R R R T L F F G G E D L R K I R Y H K V V A

2901 CATCTAAMAC TGAACAGAG AATTCTCTCT TTGCTTCTC ACTTCTCTG ATTAGAAAG AACCGCTTT AACCGCTTT TATATAGTT GTTCTCTG
 GTACATTTTG ACTTCTCTC TTACAGGAA AACCCGAAG TGAAGAGAC TAATCTTTCC TTGCGCAGAA CAAGAGCTT ATATTATCAA CAACAGACG
 159 G D L V S L S F E G K Q S E S E R I L F S G T K N E S Y L L Q K E A

3001 ATCTGATCTT CCGATTGCTC CAAGAACCA CCGCTCTGCC TGTAGGCTTC TGTCTCAGC CAGGTAGTTA GAAGGAATAT AGCTTGTAG TTGCTGACTG
 TAGACTAGAA GGCTAACGAG GTTCTCTGTT GCGCAGACGG ACATCCNAG ACAGGAGTGG GTGCATCAAT CTCTCTTATA TCGGNACATC AACGACTGAC
 125 D S H G I A G F F W P E A Q L S R D E A V Y N S P I Y G Q L Q Q S

3101 GAGCATCTC GTCTTTTCTC CAAGTGTCTG GCAACACACC AGCCCTCATG CAAGGTGTC AGAATTTGAA GTTTGTACCC TGTCTGGAG CTCAGTCTCT
 CTCGTAGAG CAGAAAGAG GTTCACAGAC CGTTTGTGTG TCGGGAGTAC GTTTCACAGG TCTTGAAGT TCAACAGTGG ACCAGCTTTC GAGTTCAGGA
 92 S G D R R K E L H R A F W W G E H L T D L V Q L K D G A R F S L D

FIGURE 5E

3201 CAGCAGTCCG AGCCTGGTAA TCAACAAG CCACAAGTA GTGCCCATGC CTCTGTGACT GGGGAGAGCA AAGGCCCTT GGATTTTCAA TCACGGTTGA
 GTCTGCAGGC TCGGACCAAT AGTTTGTTC GGTGTTTCAT CACGGGTAGC GAGACACTGA CCCCTCTCGT TTCCCGGGA CCTAAAGTT AGTCCCACT
 59 E A T R A Q Y D F L A V F Y H G H R Q S O P S C L A G P N E I V T S

 3301 CTGTCTN:CC TCCGTGACA AACAGGGAG ATAGGGTTCT AGTACTCC AGGCTCTG ACAGATGTTG CTCAATGTG CTGTGTGGG AGAAGAGGAG
 GAACAGACGG AGGCACCTGT TTGTCCCTC TATCCCAAGA TCCATGAGGG TCTCGGAGAC TCTCTACAC GAGTAACAG GAACACCCC TCTTCTCCTC
 25 K D A E T S L C P L Y P E L Y E W L R Q C I N S M

 3401 CAGGGCTTCT CCTCTCCTC TTAGTCTCTG CGATCCACCT TATCTTCTT CACCAGGCA CTTTGAAGTC AGCACCACCT CACCATACTT CGGAGAGTAT
 GTCCCAAGA GGGAGAGGG AATCAGAGAC GTTAGGTGGA ATAGAAGGA GTGTCTCGT GAACTTCAG TCGTGTGTA GTGGTATGA GCCTCTCATA

 3501 GCAAGTCCC GTTTCAGATC AGTCCAGCAG CTGGGTGCA GCAAGTCTA CTTACCGCT TCTTCTCTT GCTTCTATCT CTTCTATCT GCTTAAGNAT CCCACNACA AATAAATA AATTAAAG
 CGTTTCAGG CAAGTCTAG TCAGGTCTG GACCAACGT CGTTCAGGAT GGACCTCTCT GAATGCGGA AGAAGACA CCGACCTCA CGATCGGGCT

 3601 GCAAAACTG AGCAGGAGCT GGGCAGCTG TCACTAGGA GGTGCTTTT CTCTTATCT GCTTAAGNAT CCCACNACA AATAAATA AATTAAAG
 CCGTTTTCAG TCGTCTCTGA CCGTCTGAG AGTCTCTT CCACTAGGA GGTGCTTTT CTCTTATCT GCTTAAGNAT CCCACNACA AATAAATA AATTAAAG

 3701 GCTTTTATTT AGACAATAT CTGAGNACAG AATGCTGCA TCTTCCCTTT TGTCCCAATA AAGGTTAGC AAGAGGAGC TACTAACCCC TGGTAAACC
 CCAAAATANA TCTGTTTATA GACTCTTCTG TTACCAAGCT AGAAGGANA ACAGGCTTAT TTTCATATG TTCTCTCTG ATGATTGGG ACCATTTTGG

 3801 TCCAGTCTTT GCTTTCGCA GGTGCTGCTC GAGTATCTT CCATACCTAC CAGTCTCTG CCTGAGGTC GCGGCGGGA CTCTAGAGTC GACCTGCAGA
 AGTTCAGAA CGAAGCGGT CCCAGCTGAG CTCCCTAGNA GGTATGATG GTCAAGAGCC GACGCTCCAG CCGCGGCGCT GAGATCTCAG CTGGACGTCT

 3901 AGCTTGCGG CCATGGGCA ACTTGTTTTAT TGCAGCTTAT AATGTTTACA AATAAGCA TAGCATCACA AATTTCACA ATAAAGCAT TTTTTCACCTG
 TCAACCGG GATACCGGT TGAACAATA AGTCAATA TTACCAATGT TTATTTCTG TTAAGTGT TATTCTGTA AAAAGTAC

19/35

FIGURE 5F

4001 CATTTCTAGTT GTGGTTTCTC CAAACTCATC AATGTATCTT ATCATGTCTG GATCGGAAT TAATTCGGG CAGCACCATG GCTGAAATA ACCTCTGAAA
GTAAGATCAA CACCAACAG GTTTCAGTAG TTACATAGTA TAGTAGAGC CTAGCCCTTA ATTAAGCGG GTGCTGTAC CGACTTTAT TGGAGACTTT

4101 GAGGAACCTG GTTAGGTACC TTCTGAGGG GAAACAAACCA GCTGTGGAAT GTGTGTCAGT TAGGGTGTGG AAGTCCCCA GGCTCCCCAG CAGGCAGAA
CTCCTTGAA CAAATCCATGG AGACTCCGC CTTTCTTGGT CGACACTTA CACACAGTCA ATCCACACC TTTCAGGGT CCGAGGGGTC GTCCGTCTTC

4201 TATGCNAAC ATGCATCTCA ATTAGTCAGC AACCAGTGT GGAAAGTCCC CAGGCTCCCC ACCAGGAGA AGTATGCANA GCATGCATCT CAATTAGTCA
ATACGTTTCG TACGTAGAGT TAATCAGTCG TTGCTCCACA CCTTTCAGGG GTCCOAGGGG TCGTCCGTCT TCATACGTTT CGTACGTAGA GTTAATCAGT

4301 GCAACCATAG TCCCGCCCT AACTCCGCC ATCCCGCCC TAATCCGCC CAGTTCCGCC CATTCGCCG CCCATGGCTG ACTAATTTTT TTATTTATG
CGTTGGTATC AGGCGGGGA TTGAGGGCGG TAGGGCGGG ATTGAGCGG GTCAAGCGG GTACAGGGG GGTACCGAC TGATTAAAAA AATAAATAC

4401 CAGAGGCCGA GGCGCCCTCG GCCTCTGAGC TATTCCAGAA GTACTGAGGA GGCTTTTTC GAGGCTAGG CTTTTCANA AAGCTGTMA CAGCTTGGCA
GTCTCCGCT CCGCGCGAGC CGGAGACTCG ATAGGTCTT CATCACTCCT CCGAANAAC CTCCGANTCC GANAACGTTT TTCGACAAAT TTCGAACCGT

4501 CTGCGCGTCG TTTTACAACG TCGTGACTGG GAAACCCCTG GCGTTACCCA ACTTAATCG CTTCGAGCAG ATCCCCCTTT CCGCAGCTGG CGTAATAGCG
GACCGGCAGC AAAATGTTGC AGCACTGACC CTTTTGAGC CGCAATGGGT TGAATTAGCG GAACTGCTG TAGGGGANA GCGGTCGACC GCATTTATCGC

4601 AAGAGGCCCG CACCGATCG CCTTCCCAAC AGTTGCGCAG CCTGAATGGC GAATGGCCC TGATGCGGTA TTTCCTCCTT ACGCATCTGT GCGGTATTTC
TTCTCCGHC GTGGCTAGCG GGAAAGGTTG TCAACCGTC GCATTTACCG CTTACCGCG ACTACGCCAT AAGAGAGNA TCGGTAGACA CGCATNAAG

4701 ACACCGCATA CGTCAAGCA ACCATAGTAC GCGCCCTGTA GCGCGCATT AAGCGCGCG GTGTGCTGG TTACGCGCAG CGTGACCGCT ACACTTGCCA
TGTGGGTAT GCAGTTTCTGT TGGTATCATG CCGGAGCAT CGCCCGTAA TTGCGCGCG CCACACACC AATGCGGTC GCACCTGGCA TGTGMAACCGT

20/35

FIGURE 5G

4801 GCGCCCTAGC GCGCGCTCCT TTCCGTTTCT TCCCTTCCTT TCTCGCCACG TTGCGCGCT TTCCCGCTCA AGCTCTAAT CCGCGGCTCC CTTTAGGCTT
CGCGGATCG CCGCGAGGA AGCGAAGA AGCGGCTGC AGCGGCCGA AGCGGAGT TCGAGATTA GCCCGGAGG GAAATCCCA

4901 CCGATTTACT GCTTACGGC ACCTCGACC CAAAACCTT GATTTGGTG ATGTTTCACG TAGTGGCCA TCGCCCTGAT AGACGGTTTT TCGCCCTTTG
GGCTAAATCA CGAAATGCG TCGAGCTGG GTTTTGTGA CTAACCCAC TACCAAGTGC ATCACCCGGT AGCGGACTA TCTGCCAAM AGCGGAAC

5001 ACGTTGGAGT CCACGTTCTT TAATAGTGA CTCTGTTC CAACTGGAC AACACTCAAC CCTATCTCG GCTATTTCTT TGAITTTATA GGAITTTG
TGCACCTCA GGTGCAAGNA ATTATCACCT GAGACNAGG TTTCACCTTG TTGTGAGTTG GGATAGAGCC CGATAAGAA ACTAATATT CCTAANACG

5101 CGATTTCCGC CTATTGGTTA AAAATGAGC TGATTTAACA AAAATTTAC GCGAATTTTA ACAATATT AACGTTTACA ATTTTATGGT GCACCTCTAG
GCTAAGCCG GATAACCAAT TTTTACTCG ACTAATTTGT TTTTAAATTT CGCTTAAAT TGTTTTATA TTGCAATGT TAAATACCA CGTGAGATC

5201 TACAACTGC TCTGATGCC CATAGTTAAG CCAAGCCCGA CACCCGCCA CACCGCTCA CCGGCTGTG TGTCCCGG TGTCCCGG ATCCGCTTAC
ATGTTAGAC AGACTACGC GTATCAATTC GGTGCGGCT GTGCGGCT GTGCGGACT GCCCGGACT GCCCGACAG ACGAGGCGG TAGGCGAATG

5301 AGACNAGCTG TGACGCTC CCGGAGCTG ATGTGTCAGA GGTTTTCACC GTATCACCG AAAAGCGCA GACGAAGCG CTTCTGTATA CGCTATTTT
TCTGTTGAC ACTGCGAG GCCCTCGAG TACACAGTCT CCAAAAGTGG CAGTAGTGC TTTCGCGCT CTGCTTTCC CGAGCACTAT CGGATNAAA

5401 TATAGTTAA TGTATGATA ATAATGGTTT CTTAGAGTC AGTGGCACT TTTCGGGNA ATGTGGCGG AACCCCTATT TGTTTATTTT TCTAATACA
ATATCCAAIT ACAGTACTAT TATTACCANA GATCTGCAG TCCACCGTCA AAGCCCTT TACACGCGC TTGCGGATTA ACAATNAAA AGATTTATGT

5501 TTCANATATG TATCGCTCA TGAGACANTA ACCCTGATA ATGCTTCAAT AATATTGANA AAGGAAGAT ATGAGTATTC AACATTTCCG TGTGCGCTT
AAGTTTATAC ATAGCGAGT ACTCTGTTAT TGGACTATT TACGAAGTTA TTATACTTT TTCTTCTCA TACTCATAAG TTGTAAAGC ACAGCGGAA

21/35

FIGURE 5H

5601 ATTCCCTTTT TTGCGCATTT TTGCGTTGCT GTTTTTCCTC ACCCAGAAC CCTGGTGAAA GTAAAGATG CTGAAGATCA GTTGGGTGCA CGAGTGGGTT
TAAAGGAAAA AACGCCGTAA AACGCCGTAA AACGGAAGGA CAAAACGAG TGGTCTTTG CGACCACTTT CATTTCTAC GACTTCTAGT CAACCCACGT GCTCACCCAA

5701 ACATCCAACT GGATCTCACC AGCGGTAAAG TCCCTTCAGAG TTTTGGCCCC GAAGAACGTT TTCCAAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGGCC
TGTAGCTTGA CCTAGATTG TCGCCATTCT AGCACTCTC AAGAGCGGG CTTCCTTGCA AAGTTACTA CTGTGAAA TTTCAAGACG ATACACCGCG

5801 GGTATTATCC CGTATTGACG CCGGGCAAGA GCACCTCGGT CCGCGCATAC ACTATTCTCA GAATGACTTG GTTGAGTACT CACCAGTCAC AGAAAGCAT
CCATAATAG GCATAACTGC GCGCCGTCT CGTTGAGCCA GCGCGGTATG TCATAGAGT CTTACTGAAC CAACTCATGA GTGGTCAGTG TCTTTTCGTA

5901 CTTACGGATG GCATGACAGT AAGAGATTA TGCAGTGTG CCATAACCAT GAGTGATAAC ACTGCGGCCA ACTTACTTCT GACAACTATC GGAGACCGA
GAATGSCCTAC CGTACTGTCA TTCTCTTAAT ACGTCACGAC GGTATTGCTA CTCACTATTG TGACCGCGGT TGAATGAAGA CTGTTGCTAG CCTCCTGGCT

6001 AGGAGCTAAC CGCTTTTTTG CACAACATGG GGCATCATGT AACTCGCCTT GATCGTTGGG AACCGAGCT GAATGAAGCC ATACAAACG ACAGCGGTGA
TCCTCGATTG GCGAAAAAC GTGTTGTACC CCTAGTACA TTGAGCGGNA CTAGCAACCC TTGCGCTCGA CTTACTTCGG TATGTTTTGC TGCTCGCACT

6101 CACCACGATG CCTGTAGCAA TGGCAACAC GTTGGGCAAA CTATTAACTG GCGAATCTACT TACTCTAGCT TCCCGGCAC AATTAAATGA CTGGATGGAG
GTGGTCTAC GCACATCGTT ACCGTTGTTG CAACCGCTT GATAATTGAC CGCTTGATGA ATGAGATCGA AGGCGCGTTG TTAATTATCT GACCTACCTC

6201 CCGGATAAAG TTGCAGGACC ACTTCTGGC TCGGCCCTTC CCGCTGGCTG GTTTATGCT GATAAATCTG GAGCGGTGA GGTGGGTCT CCGCGTATCA
CGCTATTTC AACGTCTGG TGAAGACGG AGCCGGAAG GCCGACCGAC CAAATAACGA CTATTTAGAC CTGCGCACT CGCACCCAGA GCGCCATAGT

6301 TTGCAKCACT GGGGCCAGAT GGTAAAGCCCT CCGTATGCT AGTTATCTAC ACGACGGGA GTCAAGCAAC TATGATGAA CCAATAGAC AGATCGCTCA
AACGTCTGA CCGCGTCTA CCATTCGGGA GGCATAGCA TCAATAGATG TGCTGCCCT CAGTCCGTTG ATACCTACTT GCTTTATCTG TCTAGCGACT

22/35

FIGURE 51

6401 GATAGGTCC TCACTGATTA AGCATTTGGTA ACTGTGAGAC CAAGTTTACT CATATATACT TTAGATTGAT TTAAACTTC ATTTTAAAT TAAAGGATC
CTATCCACCG AGTGACTAAT TCGTAACCAT TGACAGTCTG GTTCAATGA GTATATATGA NATCTAATA AATTTGAG AATTTCTAG TTTCTCTAG

6501 TAGGTGAAGA TCCTTTTGA TAATCTCATG ACCAAATCC CTTAACGTGA GTTTTCGTTC CACTGAGGT CAGACCCCGT AGAAAGATC AAAGGATCTT
ATCCACTTCT AGGAAAACT ATTAGAGTAC TGGTTTTAGG GAATTGCACT CAAAGCAG GTGACTCGCA GTCTGGGCA TCTTTCTAG TTTCTCTAG

6601 CTTGAGATCC TTTTTCCTG CGCGTAATCT GCTGCTTGA ACAAATAA CCACCGCTAC CAGCGGTGGT TTGTTTCGG GATCAAGAGC TACCAACTCT
GAACTCTAGG AAAAAAGAC GCGCATTAGA CGACGAACGT TTGTTTTCCT GGTGGGATG GTCCGCCACA ACAAACGGC CTAGTTCTCG ATGCTTCAGA

6701 TTTTCCGAAG GTAACTGGCT TCACGAGAGC GCACATACCA AATACGTTC TTCTAGTGA GCCGTAGTTA GGCACCACT TCAAGAACTC TGTAGCACCG
AAAGGCTTC CATTGACCGA AGTCGTCTCG CGTCTATGCT TTATGACNAG AAGATCACT CGGCATCAAT CCGTGTGTA AGTTCTTGA ACATCGTGGC

6801 CCTACATACC TCCTCTCTCT AATCTCTGTTA CCAGTGGCTG CTGCGAGTGG CGATAAGTGG TGCTTTACCG GGTTCGACTC AAGACGATAG TTACCGGATA
GCATGTATGG AGCGAGACCA TTAGGACAAAT GGTCAACCGAC GACGTCACC GCTATTTCAGC ACAGATGGC CCACCTGAG TTCTCTATC AATGGCTAT

6901 AGCGGCACCG GTGCGGCTGA ACGGGGGTGT CGTGCACACA GCGCAGCTTG GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG AGCTATGAGA
TCCGCTCTGC CAGCCCGACT TGCCCCCCMA GCACGTGTGT CCGGTGMACT CTGCTTTGCT GGATGTGCT TGACTCTATG GATGTGGAC TCGATACTCT

7001 AAGCGCACCG CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG GCAGGTCTCG AACAGAGAG CGCAGGAGG AGCTTCCAGG GGAACCGCC
TTCCGCTGTC GAAGGCTTC CCTCTTTCCG CCTCTCCATA GGCCATTCGC CGTCCAGCC TTGTCTCTCTC GGTGTCTCC TCGAAGGTCC CCTTTGCGG

7101 TGTATCTTT ATAGTCTGT CCGGTTTCCG CACTCTGAC TTGAGCGTCTG ATTTTGTGA TGCTGTCA GGGGCGGAG CCTATGAAA AACGCCAGCA
ACCATAGAA TATCAGACA CCCCMAAGCG GTGCACTG AACTCGCAGC TAAACACT ACAGACATG CCCCCCTC GGATACCTTT TTGCTCTCT

23/35

FIGURE 5J

7201 ACGCGGCTT TTTACGGTTC CTGGGCTTTT GCTGGGCTTT TGCTCACATG TTCTTTCCTG CGTTATCCCC TGATTCTGTG GATAACGTA TTACCGCCTT
TGCGCCGGAA AATGCGCAG GACCGGAAAG CACCGGAAAG ACCAGTGTAC AGAAGGAC GCATAGGGG ACTAAGACAC CTATTGGCAT AATGCGGNA

7301 TTAGTGAGCT GATACCGCTC GCGGAGCCG AACGACCGAG CGCAGCGAGT CAGTGACCGA GGAAGCGGA GAGCGCCAA TAGGCNAJCC GCCTCTCCCC
ACTCACTCGA CTATGGGAG CGGCGTCGGC TTGCTGGCTC GGTGCTCA GTCACCTGCT CCTTGGCTT CTGCGGGTT ATCGTTTGG CGGAGAGGGG

7401 GCGCGTTGAC CGATTCAATTA ATGCAGCTGG CACGACAGGT TTCCGACTG GAAAGCGGC AGTGAGCGCA AGCMAATTAA TGTGAGTTAG CTCACCTCATT
CAGCGAACC GCTAAGTAAT TACGTGACC GTGCTGTCCA AAGGCTGAC CTTCGCGCG TCACTGCGT TCGTTAATT ACACTCAATC GAGTGAGTAA

7501 AGGCACCCCA GGCTTTACAC TTTATGCTTC CGCTCGTAT GTTGTTGGA ATTGTGAGCG GATAACAATT TCACACAGGA AACAGCTATG ACATGATTAC
TCCGTGGGT CCGAATGTG AATACGAAG GCGGAGCATA CAACACACCT TAACACTGCG CTATTGTTAA AGTGTGCTT TTGCGATAC TGACTAATG

7601 GAATTAA
CTTAATT

24/35

1 GCGCCCCAG ACAAGCAGA GGATGGGCT TACCAGCTCG CAGAGCCAGG ACCGGGGAGG TAGCAGAAAG ACCACAAGTA CAAAGAAGTC CTCAAACTTT
CGCCGGCGTC TCTTTGCT CTAACCCCGA ATCGTCGACC GTCGCGTCC TCGCCCCC TCCTCTTTC TGGTTCAT GTTCTTCAG GACTTGAAA

101 GGTTTTGCTG CTGCAGCCCA TTGAGAGTGA CGACATGGAG CACAAGACCC TGAACATCAC CGACTTTGGC CTGCCCCGAG AGTGGACAA AACACACAA
CCAAGACGAC GACGTGGGT AACTCTCACT GCTGTACCTC GTGTCTGGG ACTTCTAGTG GCTCAACCG GACCGGCTC TCACCGTGT TTGGTGTGT

201 ATGAGTGCCG CXGGCACCTA CXCTGGATG GCTCTGAGG TTATCAAGG CTCACCTTC TCTAAGGCA GTGACGTCTG GAGTTTGGG GTGCTGCTG
TACTCAGGC G?CCCTGGAT G?GGACCTAC CGAGCACTCC ANAGTCTCC GAGGTGAAG AGATTCCCT CACTGCAGAC CTCAAAACCC CACGACONCA

301 GGGAACTGCT GACCGGGAG XTGCCATACC GTGGCATTGA CTGCTTGCT GTGGCCTATG GCGTAGCTGT TAACAAGCTC ACATGCCAT CCATCCACCT
CCCTTGACGA CTGGCCCCC ?ACGCTATGG CACCGTAAC TACCGAACGA CACCGGATAC CCGATCGACA ATTGTTGAG TGTGACGCTA GGTAGGTGA

401 GGCC
CCGG

FIGURE 6

FIGURE 7A

1 ATGACAGCTT TGGGCGCA GGGGGGAG CCGGGCTG TGGTCTTT TTCTGCAATG ATATTGGA CTATTACAA TCAAGATGTC GCTGTGATCA
TACTCTGCA ACGGGGCT GGGGGCTC GAGGGAGG AGCAACAAA AAGAGCTTAC TATAAGCT CATAATGTTT AGTTCTAGAG GGAACACTAT

101 AGCTGCTTTT AATCAATCAT AAGACAATG ATTUATCT GGGAACTTA TATCATATC GCATGCTATC AGAATCGCG GAAGACATGTC GGTGTGCTTT
TCACACAAA TTAGTTAGTA TTTTGTTC TAACTACT GGTCTTACT ACTATATAG GCTACCATAG TCTTAGGCG CTCTGCAUC GCACAGGCA

201 GAGACGGCAG AGCTGAGGA CAGTGTAGCA AGTGTGCT GTGAGCTG AGTATCTTC TTGATCACA CTGCACTGTC TGGTCGATGC CCGAGGGAAC
CTGCGGGTC TCGATCGCT CTCACATGCT TGACGCGCA CAGTTTCAG TACATAGAG AGCTATGCT CAGCTTCAG ACCAGCTAGG GGTCCCTTG

301 ATTCTGTC TGTGCTCTT TAAGCAGC TGGCTAAT GGTAGGACA TTTTATTTA CAAGACAGC GAGTGTGTTT CATGCTCATT TTGAAATGA
TAAAGGAG AGAGGAGAA ATTGCTGTC AGGACTTAA GGTGGGCT AAACTAAT GTTTGCTC CTACACAAAG GTACCACTAA ACTTTTACT

401 CAGAAAGCA AGCTGAGAA TACTACTTT TTATTAGAG TGAAGTATC AATTACAAA TATGTGTTAC AATGATATA AGAATAGCC TCTTTAGAC
GTTTGTCT TCGACTCTT ATGATGAAA AATAAGTGC ATTGATGTC ATTGATGTC TTAATGCTT ATAAACAATG TCACTCATAT TCTTTATGCG AGCAATGTC

501 ATTAACAGA CTTTACTTTA GAAAATGCA AAGCAGGAG GGTGCTGCT GCATATCTCA GAGGCTTCCA GAGCGGATCC TGGATGGGT GCTTGGCAT
TAATTTCT GAAATCAAT GTTTTACT TTGCTGCTC GGGAGGAGA GGTATAGCT CTGCGAAGT CTGCGCTAGG ACCTTACCA CCAAGGCTA

601 TCAACGGG AGACTGTAA AGAGAAAT GTAGCTGTC TTAAGAGA GGAAGAGT CTTCATCAT TATTGGGAC GCACATAGAG TGTGTGCCA
AGTGGGCG TTGACATTT TCTTCTTCA GTGAGAAC AATTGCT CTCTTTCG GAAGTACTTA ATAAAGCTG CCGTATATCC AGACACGGT

701 GAAATGACT GGGAGGGA TCGAGAGG TTTTCAAT AGATTAAT CAAGCTGTC AGACACATTT GGCACAAITA TTCTTAAAG TAGGGAAGC
CTTACTTCA GGTGCTTT AGTGTGTC AAGAGCTTA TGTACATTTA GTTGAAGAG TGTGCTTAA GGTGTTAAT AAGCAATTTT ATGCTGTTG

FIGURE 7B

001 GTTATGATATA AGGTGGCAAG CTGTTCATCT GAACCATGGA TTGGGGTCA CTTGCAATT AGAAAACAAA GCACTCGAGG AGGGCAACTA CTTTCAGATG
GAATACTAT TCCAGGTTTC GACAACTACA CTTGTACTAT AAGCCCACT GAACCTTAA TCTTTTCTTT CGTCACTGTC TCCCGTTGAT GAAACTCTAC

901 ACTAGCTTAT CAACAACACAG AACTATGATA CGATTTCTCT TTGTTTTCT ATCATGATGCG GGAAGAACCG ACACGGGATA CTACACTTCT TCTTCTTCAA
TTATGATAA GTTGTCTTC TTGATACTAT GCTTAGACA AACCAAAACA TACTACTGAC GCTTCTTTC TGTGGCCTAT GATGACACA AGGAGAAGTT

1001 AACATGCCAG TCAATCAGCT TTGTTAGCA TGGTAAACAA GCGATTATA AATGCTAGCA ATTCAAGTCA AGATTATGAA ATTGACCAAT ATGAGAGTT
TCTAGGCTC ACTTAGTCA AACCAATGCT AGCATCTTT CTTAAATAT TTAGATCT TAACTTCACT TCTAATCACT TAACTGTTA TACTTCTCAA

1101 TTGTTTTCT CTCAGTTTA AACCTAGC ACAAAATACA TTGACTGGA CTTTCTCTCG AAAATCATTT CTTTGTGAGC AAAAGGTTCT TGATAACGGA
AACAAAACA CAGTCAAAAT TTGCTATGCG TGTTAATCT AATGCACT GGAAGAAC TTGATGAA GGAACATCG TTTTCCCAAC ACTATTGCT

1201 TACAGCATAT CCAAGTTTTC CAATCATAG CAGCAACACG GAGATATAT ATTCTATGCA GAAATCATG ATGCCCAAT TACCAAAATG TTCACCGCTGT
ATGCTGATA GCTTCAAAAG GTTAGTATTC GTGCTGCTTC CTTTATATA TAACTACT CTTTACTAC TACGGTTAA ATGTTTTAC AAGTGCACA

1301 ATATAAAGG GAAACCTCAA GTGCTGGCAG AAGCTTGCG AATGACGCG TCGTCTTCT GCGATGATA CCGATTAGCA TCTTGACCT GGAAGAAGTG
TATATCTTC CTTTGAATTT CAGGACCTC TTGCAAGCGG TTGACTGCG AGAACAACA GCTACTAT GGTAAATGCT AGAAGCTGHA CTTTCTTAC

1401 TTCAACAGG TCTGCCAAT GACACAGCA GATACACAA GAGTTTGA ATGAAAGCG TAACAGAAA CTTTTHGAC AGTGCTGTC GAGCACTACT
AAGCTGTC AGAGGTTGA CTTGCTTCT CTAGCTTCT CTTCAAGCT TATCTTTCG ATTCTCTT CACAACCTG TCACCCACAG CTGCTCATGA

1501 CTAAACATCA GTGAGCCCAT AAAGGCTTC CTGTTCAAT GTGTGATA CAATTCCTT GGCACATCT GTGAGAGCAT CTTTTAAAC TCTCCAGGCG
GATTCTACT CACTTGGTA TTTTCCCAAG GAGCACTCA GCACTATAT GTTAGGAAA CCGTGTGAA CACTCTGTA GGAATTTG AGAGTCCCG

27/35

FIGURE 7C

1601 CTTTCTCTTT CATCCAGAC AGCATCTTAT TTTATGTAAG AATTGCTTT TTTCTCTCT TATTCTCTT TTAACCAAA ACAGACGACA ATTAACAGCA AATTGGAC GATTAAACAG TTTTCATCTT
 GAAGGCAAA GTAGCTTCG TTTATAGTA ACATATCTTG TTAACCAAA ACAGACGACA ATTAACAGCA AATTGGAC GATTAAACAG TTTTCATCTT

1701 AAGCAATTT AGGTATGAA GGCATCTACA GATGGTACAG GTGACGGAT CCTCAGATAA TGAGTACTTC-TACGTTGATT, TCAGAGAATA-TUATATGAT
 TTTGCTTAAA TCCATCTTT GGTGCTATCT CTTAGCTATCT CTTAGCTATCT CTTAGCTATCT CTTAGCTATCT CTTAGCTATCT CTTAGCTATCT CTTAGCTATCT

1801 CTCAAAATCGG AGTTTCCAAG AGAAATTTTA GAGTTTGGGA AGGTACTAGG ATCAGGTGCT TTTGGAAA TGATGAACGG AACAGCTTAT GCAATTAGCA
 GAGTTTACCC TCAGAGCTTC TCTTTTAAAT CTCAAACCT TCCATGATCC TACTCCACCA AACCTTTTC ACTACTTGG TTGTGCAATA CCTTAATCGT

1901 AAACAGCAGT-CTCAATCCAG GTTACCGTCA AATGCTGAA AGAAGAGCA CACAGCTCG AAGAGAGCG-ACTCATGTCA GAATCAAGA-TGATGACCCB
 TTTGTCTCTCA GAGTTAGGTC CAATGCCAGT TTTAGCAGT TCTTTTCTGT TCTTTTCTGT TCTTTTCTGT TCTTTTCTGT TCTTTTCTGT TCTTTTCTGT

2001 GCTGGGAGCG-CACGAGAATA TTGTGACCT-GCTGGGCGG TGCACACTGT CAGGACCAAT TTACTTCTAT TTTGATATCT GTTGGTATGG TGATCTTCTG
 CGAGCTTCTG GTGCTCTTAT AACACTTGA CAGCTTCTG CAGCTTCTG CAGCTTCTG CAGCTTCTG CAGCTTCTG CAGCTTCTG CAGCTTCTG

2101 AACTATCTAA-GAAGTAAG AGAAATTT CACAGACTT GCACAGACTT GCACAGACTT GCACAGACTT GCACAGACTT GCACAGACTT GCACAGACTT GCACAGACTT
 TTGATAGATT CTTCATTTTC TCTTTTAAA GTGCTCTGA CTTGCTCTGA CTTGCTCTGA CTTGCTCTGA CTTGCTCTGA CTTGCTCTGA

2201 ATTCCAGCAT-GCTGGGTTCA AGAGAGTTT-AGATACACCG GGATCTGAT CAAATCTCAG GGTCTCATGG CAATTCATTT CACTCTGAG ATCAATTTGA
 TAAGGTGCTA CGGACCAAGT TCTCTTCAG TCTATCTGG CTTGAGCTA GTTACAGTC CGGAGTACC CTTAAGTANA GTGAGACTTC TACTTAACT

2301 ATATGAAGC CAAAAAGCG TGAAGACA GAGGAGCTTG AATGCTGTA CATTGAGA TCTCTTTCG TTTGATATC AAGTGCCTAA AGAATGCA
 TATCTTTG GTTTTTCG ACCTTCTCT CTTGCTGAG TTACAGCAAT GTAACTTCT ACAGAGAGCG AACGTATAG TTCAGCGGT TCTTACCTT

28/35

FIGURE 7D

2401 TTTCTGGAAAT TTAAGTCGTG TGTTCACAGA CACCTGCGCG CAGGAAAGGT GCTTGTACCC CACGGGAAAG TGGTGAAGAT ATGTGACCTT GATTGGCTC
 A AAGACCTTA AATTCAGCAC ACAAGTGCT CTGACCGCG GTTCCTTCCA CCAACAGTGG GTCCCTTTC ACCACTTCTA TACACTGAAA CCTAACCGAG

2501 CAGATATCAT GAGTGATTCC AACTATGTTG TCAGGGGCAA TCCCGTCTG CCTGTAAAT GATGCGCCC CGAAAGCTG TTTGAGGCA TCTACACCAT
 CTCTATGCTA CTCACTAAGG TTGATACAA TTTTCCGTT ACAGGACAC GGACATTTTA CTACCGCGG GCTTTCGGAC AAATCTCGT AGATGTGTA

2601 TAAGATGAT GTCTGTGAT ATGGATATT ACTGTGGA ATCTTCTCAC TTGTGTGA TCTTACCTT GGCATTCCG TGTATGCTAA CTCTACAAA
 ATTCTCACTA CAGACCACTA TACCTTATAA TCACACCTT TAAAGATG AACACACT AGCAATGGA CCGTAAGGCC AACTACGATT GAAGATGTTT

2701 CTGATTCAAA ATGATTTAA AATGATCAG GCATTTTATG CTACAGAAAG AATATACAT ATAATGCAAT CCGCTGCGG TTTTACTCA AGCAAGCGG
 GACTAATTT TACCTAATTT TTACCTAGTC GTTAAATATC CATCTCTCT TTATATCTAA TATTACCTTA GGACGACCGG AAACTGACT TCCCTTCCG

2801 CATCCCTCCG TAAATTTGACT TCGTTTTAG GATGTACGCT GGCAGATCCA CAAGAAGCA TGTATCAGAA TGTGATGCG CCGTGTTCG AATGTCTCA
 GTAGGAAGG ATTAAGTGA AGCAAAATC CTACAGTCCA CCGTCTAGT CTCTTCGCT ACATAGTCT ACACCTACCG GCACAAAGCC TTACAGCAGT

2901 CACCTAGCAA AACAGCGGAC CTTTACGAG AGAGATGAT TTGGGCTAC TCTCTCCGA GCGTCAGGTC GAAGATTGCT AGAGAACAA TTTAGTTTTA
 CTGGATGCTT TTGTCCGCTG GAAAGTCGT TCTCTACCTA AACCCGATG ACAGAGCGGT CCGAGTCCAG CTTCTAAGCA TCTCTTGT AAATCAAAAT

3001 AGGACTTTCAT CCGTCCACCT ATCCCTAACA GGCTGTAGAT TACCAAAACA AGTTAATTT CATCACTAAA AGAAATCTA TTAACAACG CTGCTTACG
 TCTCAAGTA GGGAGGTGGA TAGGATTTGT CCGACATCTA ATGGTTTTGT TCCAATTAAG TAGTGATTT TCTTTAGAT AATAGTGCAC GACGAAGTGG

3101 AGACTTTTCT CTACAGAGG
 TCTCAAAAGA CATCTCTCG

29/35

FIGURE 8A

1 TCGGCTCCA CCGCCGAGG GAGAGTCAGA CTTGGGGGG GAGGGGCCC CCAAGTCAG TTGGATCTT ACCGAGTGA GCGGGGGCCA TTGAGCTCCG
 AGTTCAGCT GGGGGGCTC CTCTGAGTCT GAGGCGGCG GTTCCGGGG GTTTGAGTC AGGCTAGGA TGGCTCACT CCGCCGGGT AUCTGAGGC
 M E L R

101 GTTGTCTCTC TCGTGGGCTT GGTGGGCGG AGCTTTGGA GAGAGCTGC TCAGACAAA ATTGGAACT GCTGATCTGA AGTGGTCA ATTGCTCAG
 CTACGACGAG ACCAGCCGA GCAACGCGG TCGAACTT CTCTGGAGG ACTTGTTT TAACCTTTGA CGACTAGCT TCACCCACTG TAAGGAGTC
 5 V L L C W A S L A A A I E E T L L N T K L E T A D L K W V T F P Q

201 GTGACGGCC AGTGGAGCA ACTGAGGCC CTGGATGAG GACGACAG GTGTCTGACC TACGAAGTCT GTGAGCTGCA GGTGCGCCG GACGAGGCC
 CAGCTGCGG TCACCTCTCT TCACTGCGG GAGCTACTTC TTCTGTCTC GACGCTGCG ATCTTCACA CACTGCACT GCGACGGGG CCGTCCGGG
 30 V D G Q W E E L S G L D E F O H S V R T Y E V C D V Q R A P G Q A H

301 ACTGGCTTGG CACAGTTGG GTCCACGCG GGGGTTTGT CCAGCTGTAC GCGAGCTGCG GTTTCACCAT GCTGAGTGC CTGTCCCTGC CTGGGCTGG
 TGACCAAGC GTGTCCACC CAGGTTTGG CCGCTGCGA GTTGCACATG CCGTGGCAGG CGAGTGGTA CGAGCTCAG GACAGGAG GAGCCGACC
 72 W I R T G W V P R R G A V H V Y A T L R F T M L F C L S L P R A G

401 GCGCTCTGCG AAGGAGACT TCACCTCTT CTACTATAG AGGATGCGG ACAGGCGAG GCGCTCAGG CCAGCTGGA TGGAGNACC CTACATCAAG
 CCGGAGAGG TTCTCTGGA AGTGGAGAA GATGATATC TTCTACGCC TGTCCGCTG CCGGAGTGC GGTGAGACT ACCTCTTGG GATGTAGTTC
 105 R S C K E T F T V F Y Y E S D A D T A T A L T P A W H E N P Y I K

501 GTGGATACGG TGGCGCGGA GCATCTACC CGGAGCGCC CTGGGCGGA GCGCAGCGG AAGTGAATG TCAAGAGCT GGTCTGGA CCGCTCAGCA
 CACCTTACC ACCGCGCT COTAGAGTGG GCTTTCGGG GACCCCGCT CCGTGGGCC TTGCACTTAC AGTCTGGA CCGAGACCT GCGAGTCT
 130 V D T V A A E H L T R K R P G A E A T O K V N V K T L R L O P L S K

601 AGCTCTGCTT CTACCTGCC TTCCAGACC AGCTTCTCTG CATGGCCCTG CTATCCCTGC ACCTCTTCTA CAAGAGTGC GCGCAGCTGA CTGTCAACT
 TCCGATCTGA GATGAGCGG AGCTCTCTGG TCCGAGGAC GTACCGGAC GATAGGAGG TCGAGAGAT GTTTTCAG CCGTGGACT GACACTTGA
 172 A G F Y L A F O D O G A C H A L L S L H L F Y K K C A Q L T V N L

FIGURE 8B

801 TACTGGCTG AGGATGGCA GTGGGGCA CAGGCTCA GGGCTTCG CTGTGCTCG GAGTTCAG CAGCTGAGG GAACCAAG TCCCGAGCT
 ATAGGACAC TCTACGCT CACCGCTT GAGGCAAT GGGGAGTC GAGAGAGC CCAAGTTC GTGACTCC CTGTGCTC AGGCTCGA
 210 Y C R E D G Q W A E Q P V T G C S C A P G F E A A B G N T K C R A C

 901 GTGTGAGG CAGCTTCAAG GGGCTTCAG GAGAGGTC GTGGAGCA TCCCGAGCA TACCCATC TAACCATT GATCAGCG TCTCCAGTG
 CAGGCTCC GTGGAGTTC GGGAGATC GTTTCAG GAGTCTGT AGGCTGCT ATGTGAG ATTGTGTA CTTAGTGG AGACGTCAC
 272 A Q G T F K P L S G E G S C Q P C P A N S H S N T I G S A V C Q C

 1001 CCGCTGCG TACTTCGG CAGGACAGA CCGCGCTGT GAGCTTCA CCACTCTCC TTGGCTCG GAGGCTGG TTTCGGCT GAAGCTCC
 GGGGAGCC ATGAGGCCC GTGCTGTCT GAGGCTCA CTTGGAGCT GTGGGAGG AGCGGAGC GCCTGCAC AAGGCGCA CTGGCGAGG
 305 R V G Y F R A R T D P R G A P C T T P P S A P R S V V S R L N G S

 1101 TCCCTGACC TGGATGGAG TCCCGCTG GAGTGTGT GGGAGAGA CTTCACTAC GCTTCGCT GCGGAGTG CCGAGCGGA GCTCCTGTG
 AGGAGCTG ACCTTACCT CCGGGGAC CTCAGGAC GGTCTCTCT GAGTGTAT GGGAGAGA CCGCTCTAC GCTGGCTC CCGAGCAC
 338 S L H L E W S A P L E S G G R E D L T Y A L R C R E C R P G G S C A

 1201 CCGCTGCG GGGAGCTG ACTTTGAG CCGCGCTCG GAGCTGCT GAGGCTGG TGTGTCTG AGGCTAGT CTTGACTTCA CTTATACCT
 372 P C G G D L T F D P G P R D L V E P W V V V R G L R P D F T Y T F

 1301 TGAGTCACT GCATTGAG GGTATCTC CTTAGGAG GGGCGCTC CATTTAGCC TGTCAATG ACCACTGAC GAGAGTACC TCCGCACTG
 ACTCACTGA CTTACTTC CCAATAGG GAATGCTG CCGGGAGG GTAACTCG ACAGTTACG TGTGACTG CTTCCATGG AGGAGTCAC
 405 E V T A L N G V S S L A T G P V P F E P V N V T T D R E V P P A V

 1401 TCTGACAT GGTGACCG GTCTACCC AGGACTTCA GCTGCTG GCTGTCTC CCGGACCA GTGGGCTGT GCTGACTAC GAGTCAAT
 AGACTTAG CCACTTCC CAGGATGAG TGTCTCACT CACCGGAC CCGAGAGG GCTGTGCT CACCGACA CCACTGATG CTCAGTTTA
 430 S D I R V T R S S P S S L S L A W A V P R A P S G A V L D Y E V K Y

31/35

FIGURE 8C

1501 ACCATCAGAA GGGGGGAG GATCCAGCA GGTGGTGT CTTAGAGG TCAGAAACG GCGCAGCT GCGGGGCTG AACGGGAG CCACTACT
 TGTACTTT CCGGGGCTC CAGAGTGT GATATTA GACTTTC AGTTTTCG CCGTTCGA CCGGGGCGAC TTGGGGCTC GGTGATGA
 472 H E K G A E G P S S V R F L K T S E N R A E L R G L K R G A S Y L

 1601 GHTGAGTA GGGGGGCT CTAAGCGCG CTAGGCGG TTGGGCGG AACATACAG CCAGACCCA CTGATCAGA GCGAGGCTG GCGAGAGCA
 CCACTCCAT GCGGGCGA GACTCGGC GATGCGG AGCGCTCC TTGTATGTC GGTCTGGTT GACTACTCT GGTCCGAC GCGCTGTC
 505 V Q V R A R S E A G Y G P F G Q E H H S Q T O L D E S E O W R E Q

 1701 CTGGCCCTGA TTGGGGCAG GGCAGTCTG GATGCTGC TTCTCTGT GTCTATGTC GTCCAGTTC TCTGCTCAG GAGCGAGC AATGAGAG
 GACCGACT AACGGCGTG CCGTCAGC CCACACGAG ATCAGACCA CCAGTACAC CAGCTCAAG AGACGAGTC CTGCTCTCG TTACCTCTC
 538 I A L I A G T A V V G V V L V L V V I V V A V L C L R K Q S N G R E

 1801 AACGAGAATA TTGGACAAA CACGACAG ATCTCATCG ACATGCTACT AGCTCTACA TGGACCTCT CACTATGAA GACCTAATG AGCTCTGAG
 TTCTCTTAT AGCTGTGT GTGCTGTA TACATGAC TTACACGTA TTCCAGTGT AGCTGGGAA GTGAATACT TTGGATTTAC TCGACACT
 572 A E Y S D K H G Q Y L I G H G T K V Y I D P F T Y E D P N E A V R

 1901 GGAATTGCA AAGAGATCG ATGCTCTCTA GGTCAAGAT GAAGAGTGA TTGTGCGAG TGAATTTGC GAGTGTGCC GGGGGGCTC CAGGGCCCA
 CTTAAAGT TTCTCTAGC TACAGAGAT GAGTTCTAA CTCTCTCAT AACGCTGC ACTCAACCG CTCCACACG CCGGGGCGA GTTCCGGGT
 605 E F A K E I D V S Y V K I E E V I G A G E F G E V C R O R L K A P

 2001 GGAAGAGAG AGAGTGTGT GGAATCAAG ACCGCAAG GTGCTACAG GAGGGCAG CGAGTGTGT TTCTGAGCA GCGCTCCATC ATGCGCCAGT
 CCGTCTTCC TTCTCTAGC TACAGAGAT GAGTTCTAA CTCTCTCAT AACGCTGC ACTCAACCG CTCCACACG CCGGGGCGA GTTCCGGGT
 638 G K K E S C V A I K T L K G G Y T E R Q R R E F L S E A S I M G Q F

 2101 TCGACACCC CAATATATC GCGCTGAG GCGTGGTC CACAGCATG CCGCTCATCA TTCTCAGA GTTCTATGAG AACGGGCGC TACTCTCT
 AGCTGTGAG GTTATAGTAG GCGGACCTCC GCGACCATG GTTGTGTAC GCGAGTACT AAGAGTGT CAAGTACTC TTGGGGCGG ACCTAGCA
 672 E H P N I I R L E G V V T N S M P V M I L T E F M E N G A L D S F

32/35

FIGURE 8D

2201 CTTGCGCTA AAGCAGGAC AGTTCACAGT CATCAGCTC CTGGCAGTC TCGCGGCAT CCGCTCGAC ATGATAC ACCGAGAT GAGTACTTC
 GAGCGCGAT TTCTGCTTG TCAAGTGTCA GTAGTGTGAG CAGCGTAGG AGCGCCGTA GCGAGCTCG TAGCCATAG AACGCTCTA CTCGATGAG
 705 L R L N D G Q F T V I Q L V G H L R G I A S G M R Y L A B M S Y V

2301 CAGCGAGCC TGGCTGCTG CACATCTTA GTACAGCA ACTCGCTG CAACTGCT GACTTGGCC TTCCCGATT CCGAGGAG AACCTTCCG
 GTGGTCTGG ACCGAGGAG GTTGAGGAT CAGTTTCTGT TCGAGCAG GTTTCAGAG CTGAAACCGG AAGGCTTAA GACCTCTTC TTGAGAGGC
 738 H R D L A A R N I L V N S N L V C K V S D F G L S R F L E E N S S D

2401 ATCCGACTA CAGCAGCTCC CTGGAGGAA AGATTCCCAT CCGATGAGT GCGCGGAGG CCAATTGCTT CCGAAGTTC ACTTCGCGA GTATGCTCG
 TAGATTCAT GTGCTGAGG GACCTCTCT TCTAAGGTA GGTACCTGA CCGGCGCTCC GGTAAAGCA GGCCTTCAAG TGAAGCGGT CACTACGAG
 772 P T Y T S S L G G K I P I R W T A P E A I A F R K F T S A S D A W

2501 GAGTTACGG ATTGTGATG GCGAGTGAT GTGATTGG GAGAGCGCT ACTGGAGAT GAGCATCAG GAGGTATCA ATGCCATTGA ACAGGACTAC
 CTCATGCC TAACACTACA CCGTCACCTA CAGTAACCC CTGCGCGCA TGACCTCTTA CTGCTTAGT CTGCTAGT TAGGTAACT TGTCTGATG
 805 S Y G I V H W E V H S F G E R P Y W D M S N Q D V I N A I E Q D Y

2601 CCGTTCGCC CCGCCGACA CTGTCGCCAG TCGTCGCCAG AGTCATGCT GACTTGTGG CAGAAAGCC GGAATGCCG GCGCGCTTC CCGCAGTGG
 GCGCAGCGG CCGGCGCT GACAGGCTG AGCGAGTGG TCAATACA CCGACACCC GTCTTCTGG CCTTACGCG CCGGCGGAG GCGTCCACC
 838 R L P P P D C P T S L H Q I H L D C W Q K D R N A R P R F P Q V V

2701 TCAGCGCCT GACACAGATG ATCCGGAACC CCGCAGCCT CAAATGCTG GCGCGGAGA ATGCGCGCG CACACACCT CTCGTGACC ACCGCGGCC
 ACTTGGGGA CCGTCTTAC TAGCCTTGG GCGCTGGA GTTTAGCAC CCGCGCTCT TAGCGCGCT TAGCTGGA GAGGACCTG TCGCGCTCG
 872 S A L D K H I R N P A S L K I V A R E N G G A S H P L L D Q R Q P

2801 TCACTACTCA GCTTTGCTCT CTGCGGCGA GTGCTTCCG GCAATCAAA TCGAAGATA CGAAGAAAT TTCCAGCGG CTCGCTTTGG CTCCTTCGAG
 ACTTATCAGT CGAAGCTGA GACACCGCT CACCAACCC CCGTAGTTTT ACCCTTCTAT GCTTCTTCA AAGCTCGCG CACCGAACC GAGGAAGCTC
 905 H Y S A F G S V G E W L R A I K H G R Y E S F A A A G F G S F E

FIGURE 8E

2901 CTGCTCAGCC AGATCTCTGC TCAGCAGCTG CTCGAATCG GATCACTCT GCGCGACAC CAGAGAAA TCTTGGCCAG TGTCCAGCAC ATGAGTCCC
(ACCCAGTCCG TCTAGAGACG ACTCTTGAC GAGCTTAGC CTGCTTAGC CCGCCCTGTG GTCTCTTTT AGAACCGCTC ACAGGTGCTG TACTTCAGG
918 L V S Q I S A E D I L R I G V T L A G H Q K K I L A S V Q H H K S Q

3001 AGCCCAAGCC GCGNACCCCG GGTGGGACAG GAGCAGCGGC CCGCGAGTAC TCACCTGCG TCACTCCCA CCCCAGGAC ACCGCTTCCC CATTTCGCG
TCTGGTTCCG CCTTGGGCG CCACCTCTC CTCTGKCG GCGCTCATC ACTGAGCTC CTGAGGGT GCGTCCCTG TCGCGAGCG GTAAAGGCC
972 A K P G T P G G T G G P A P Q Y O P A G T P H P R D T A S P F S G

3101 GTCAGAGTGG GGACTCACAG AGGCCCCAG CCTGTGCC CCTGTGATTG CACTTTGACC CCGTGGGTG AGAGTTGGC AATTGGAGA GACAGATT
CCTCTCACCC CCGTGGGTC TCCGGGCTC GGCACCGG GGCACCTAAC GTCAACTCG GGCACCCAC TCCTCAACG TTAACCTCT CTGCTTAA
1005 A E W G L T E A P S P V P R W I A L O A R G V R S W Q F O E T G F

3201 GCGGTTCTG CCATAATAGG AGCGAAT CACCCCCAG CCACCTGGG GNACTCAGA CCAAGGTGA GCGGCGCTTT CCTCAGCAC TGGTGTGAC
CCCCCAAGAC GGTATTATCC TCCCTTTTA GTGGGGTCT GTTGGGCTC CTGAGCTCT GTTTCCTCT CCGCGGAAA GGGAGTCTG ACCCACACTG
1018 G G S A I I G G E N H P P A T S G N S R P R V R A P F P Q D W V O P

3301 CAGAGAAA GGAAGTCCC AACATCTCC AGCTTCCCA GTGCCCCCG TCACCTTCA TGGTGGCTC CCGCAGACCA AAGAGAGTGT GACTCCCTTG
GTCTCTTTT CCTTCACGGG TTGTAGAGG TCGAGGGGT CCAGCGGGG AGTGAACCTA CCGCCCAAG GCGTCTGCT TTCTCTCACA CTGAGGGAAC
1072 E E K E V P N I S Q P P Q V P P S P O W V R S R R P K R V O L P C

3401 CCAGTCCAG AGTGGGGGG CTGTCCAGG GCGCAGAGAG GGTGTGAGG GCCCAGTGC AAATCATTG GGTGTTGTAG TCCCACTTG CTGCTGCAC
GGTGCAGTTC TCACCCCCC GACAGGTTC CCGTTCTTC CCGACAGTCC CCGTCTACTG TTTTAGTAAC CCGAACATC AGGTTGCAAC GACGACAGT
1105 Q L Q S G G A V P G G K K G C Q G P V T K S L G F V V P T C C C H

1501 CAGCAAACTC AATCATTTTT TTCCTTGA AATGCCCTC CCCCAGTGC TCCTTCATA TCGAGCTTT TCGAGTTTG TTTTGGCT TAAATTTCT
GTGTTTGAG TTAGTAANA AAGCAACAT TTACGGGAG GGGTCCAGG ACCGATAT AACTTCAAA AACTCAAAAC AAAAACCAAA ATTAAAGA
1118 H O T O S F F S L V N A P P P A A A F I L K V F E F C P W S O F F S

34/35

FIGURE 8F

```
3601 CCGCGTTCCC TTTTGTTC TTGTTTGT TTTCTACG TCTTGTCAT AACTTGTGT TGGAGGAGC CTGTTTCACT ATGGCTCTCT TTGCCCAGT
      GGGCAAGG AAAAAAAG AAGCAAA AAGATATCC AGCAACGTA TTGAACACA ACCTGCTTG GACAAAGTA TACCGAGGA AACGGTTCA
1172 P F P F C F F V L F F Y R P C H N F V I, E G T C P T M A S F A Q V

3701 TGAACAGGG GCGCATCATC ATGTCTTTT CCAACACGT GCTTGCTCA TCCACATCC CCGACCCCG CCTGGGACC CCAGCTGTG TCTATGAAG
      ACTTTTCCC CGGTAGTAG TACAGCAAA GGTCTTGTA CCAACACGT AGGTGTAGG GCGCTGGGC GGACCTTGG GTTCCGACAC AGGATACTTC
1205 E T G A H H V C F Q N S A L V I P H P R T P P G T P K L C P H K

3801 GCGTGGGG TCAGGTAGTG AAAGGGGG TACTTGCTG TGAACCCAG AACCGACC CCGTCTTGG AGGGTTCTT AAATTATTT TAAAAAGTA
      CCCACACCC ACTCCATC TTTTCCGCC ATCAACACC ACCTTGCTC TTGCTGCG GCGACGACC TCCCCAAGAA TTTATATAA ATTTTTCAT
1218 G C G V R O O K G R O L V V E P R N G R R C L E G F L N Y I O K S N

3901 ACTTTTGT TAAATAAAG AAATGGGAC GGTCCCGC TCCAGGGTA AAAAAAAA AAAAAAAA
      TGAACACAT ATTTATTTT TTTTACCCCTG CACAGGCTG AGTCCCAT TTTTTTTT TTTTTTTT
1272 F L Y K O K K H G R V P A P G V K K K K K R
```

35/35

FIGURE 9 ARNVLVNSNLVCKVSDFGLSRFLEDDTSDPTYTSALGKIPMRWTAPETAIQYRKFFASAS

FIGURE 10 NVLVKSPNHVKITDFGLARLLEGEKEYNADGGKMPIKWMALECIHYRKFTHQ

FIGURE 11 NCMLAGDMTVCVADFGLSWKIYSGATIVRGCAKLPVKWLALGSLADNLYTVHS

FIGURE 12 NCLVGKNYTIKIADFGMSRNLYSGDYY

FIGURE 13 TRNVLVENENRVKIGDFGLTKVLPQDKKEYYKVKPEGESPIFWYAPESLTFVSD

FIGURE 14 ARNVLVNSNLVCKVSDFGMSRVLEDDPEAAVTRGGKIPIRWTAPETAIYRKFTSASD

I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all) ⁶		
According to International Patent Classification (IPC) or to both National Classification and IPC		
Int.Cl. 5 C12N15/12; C12N15/11	C12N15/54;	C12N9/12; //C12Q1/68,
II. FIELDS SEARCHED		
Minimum Documentation Searched ⁷		
Classification System	Classification Symbols	
Int.Cl. 5	C12N ; C12Q ; C07K	
Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched ⁸		
III. DOCUMENTS CONSIDERED TO BE RELEVANT⁹		
Category ¹⁰	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
A	<p>NEURON vol. 6, no. 5, May 1991, pages 691 - 704 LAI, C. & LEMKE, G. 'An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system' see the whole document</p> <p style="text-align: center;">---</p> <p style="text-align: right;">-/--</p>	1-7
<p>¹⁰ Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"A" document member of the same patent family</p>		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search		Date of Mailing of this International Search Report
07 JUNE 1993		02 -07- 1993
International Searching Authority		Signature of Authorized Officer
EUROPEAN PATENT OFFICE		ANDRES S.M.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	<p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA vol. 88, no. 23, 1 December 1991, WASHINGTON US pages 10411 - 10415 HOLTRICH, U. ET AL. 'Two additional protein-tyrosine kinases expressed in human lung: fourth member of the fibroblast growth factor receptor family and an intracellular protein-tyrosine kinase'</p> <p>---</p>	1-7, 13-15
A	<p>GENE vol. 110, no. 2, 15 January 1992, AMSTERDAM NL pages 205 - 211 BRÄUNINGER, A. ET AL. 'Isolation and characterization of a human gene that encodes a new subclass of protein tyrosine kinases' see the whole document</p> <p>---</p>	1-7, 13-15
A	<p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA vol. 87, November 1990, WASHINGTON US pages 8913 - 8917 PARTANEN, J. ET AL. 'Putative tyrosine kinases expressed in K-562 human leukemia cells' see especially clone JTK10</p> <p>---</p>	1-7
A	<p>MOLECULAR AND CELLULAR BIOLOGY vol. 9, no. 4, April 1989, WASHINGTON US pages 1587 - 1593 HAO, Q.-L. ET AL. 'Isolation and sequence analysis of a novel human tyrosine kinase gene' see the whole document</p> <p>---</p>	1-7, 16-18
P,A	<p>WO,A,9 214 748 (AMERICAN CYANAMID COMPANY) 3 September 1992 see the whole document</p> <p>-----</p>	1-12